

**The real-time molecular characterisation of human
brain tumours during surgery using Rapid
Evaporative Ionization Mass Spectrometry and
Raman Spectroscopy:
A platform for Precision Medicine in Neurosurgery.**

**Mr Babar Vaqas
BM BCh (Oxon) BSc (Hons) FRCS (SN)**

Department of Medicine, Imperial College London

Supervised by:

**Mr Kevin O'Neill MBBS BSc FRCS (SN)
Dr Nelofar Syed PhD
Professor Zoltan Takats PhD**

Submitted for the degree of Doctor of Philosophy

September 2018

Dedication

My Mother for her unwavering support, my father for his encouragement and can-do attitude, my wife Aisha for her patience and wisdom and my children Maryam, Aaminah and Yusuf who remain the energy that drives me through the impossible and to my sister Naisha for helping out just in time.

Declaration of Originality

I herewith certify that all material in this dissertation which is not my own work has been properly acknowledged.

Dr Babar Vaqas

Copyright Declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

“Following the light of the sun, we will leave the Old World”

~ Inscription on Columbus' ships ~

Similar to Christopher Columbus' voyage in 1492 which led to the widespread discovery of the New World, brain tumour surgeons have long been trying to move beyond current techniques to discover new ways of achieving their impossible task.

The task of the neurosurgeon is to remove tumours growing within the brain as completely as possible but without damaging the structure and function of the brain. The task described this way is physical, and yet it is much more than that. Tumour tissue has to be wrestled out of the same physical substance that creates our very thoughts, houses our emotions, controls our ability to speak and understand speech but also voluntarily move the entire body. The cost of failure is devastating: not only must the patient live with the resulting loss of function but further life prolonging chemo-radiotherapy treatment may be withheld if the patient is not fit enough to undergo its gruelling regime. Leaving tumour tissue behind is known to adversely affect the survival of the patient.

It is in between these paradoxical demands that neurosurgeons operate every day. During my training in neurosurgery I was painfully aware of the lack of guidance the surgeon has to fine tune this balance between preservation and destruction of brain tissue, and how all too often disability is accepted with the words “that is the best we could do”. But is it really the best that is possible?

This is the Old World. I hope to show in this thesis for the first time the New World that my research has glimpsed. This is a world where the surgeon has at their fingertips instant, real time molecular information generated during surgery giving the ability to unequivocally know what lies in the operative field. This will enable the surgeon for the first time to make decisions based on exactly where diseased tissue lies. It is also befitting that of the two technologies developed one is optically based- we are, in Columbus' words, literally following the light into a new dawn of precision medicine applied to surgery.

It is my belief that the next 20 years will see enormous leaps as neurosurgery moves from organ level precision to a more precise molecular level of precision, opening up the next chapter of one of the most exciting surgical specialities of which I have a privilege to be a practitioner.

Acknowledgements

I am indebted to my supervisors, colleagues, friends and family who have supported me through this PhD. In particular I would like to thank the Brain Tumour Research Campaign and Brain Tumour Research who have funded me through this research.

Mr Kevin O'Neill, head of Neuro-Oncology at Imperial College Healthcare NHS Trust, gave me the mentorship, time and space to innovate in a busy department. Professor Zoltan Takats and Dr Nelofar Syed provided expert supervision and guidance despite extremely busy schedules. Professor Sergei Kazarian provided expert guidance on Raman spectroscopy and become a partner in arms to introduce it to clinical practice and the wider spectroscopy community. In addition, I would like to thank the following groups of people who made this work possible:

The patients who unselfishly consented to this research and propelled me to finish it not for their benefit but for patients to come.

The clinical team, scrub nurses, pathology team and spectroscopy technicians at Charing Cross Hospital.

The academic team at the John Fulcher Molecular Neuro-Oncology Laboratory: in particular Dr Combiz Khozoie.

The British Columbia Cancer Research Agency for providing Raman equipment.

Waters Corporation for providing iKnife equipment.

Abstract

Aim

To investigate new methods for the chemical detection of tumour tissue during neurosurgery.

Rationale

Surgeons operating on brain tumours currently lack the ability to directly and immediately assess the presence of tumour tissue to help guide resection. Through developing a first in human application of new technology we hope to demonstrate the proof of concept that chemical detection of tumour tissue is possible. It will be further demonstrated that information can be obtained to potentially aid treatment decisions. This new technology could, therefore, become a platform for more effective surgery and introducing precision medicine to Neurosurgery.

Methods

Molecular analysis was performed using Raman spectroscopy and Rapid Evaporative Ionization Mass Spectrometry (REIMS). These systems were first developed for use in brain surgery. A single centre prospective observational study of both modalities was designed involving a total of 75 patients undergoing craniotomy and resection of a range of brain tumours. A neuronavigation system was used to register spectral readings in 3D space. Precise intraoperative readings from different tumour zones were taken and compared to matched core biopsy samples verified by routine histopathology.

Results

Multivariate statistics including PCA/LDA analysis was used to analyse the spectra obtained and compare these to the histological data. The systems identified normal versus tumour tissue, tumour grade, tumour type, tumour density and tissue status of key markers of gliomagenesis.

Conclusions

The work in this thesis provides proof of concept that useful real time intraoperative spectroscopy is possible. It can integrate well with the current operating room setup to provide key information which could potentially enhance surgical safety and effectiveness in increasing extent of resection. The ability to group tissue samples with respect to genomic data opens up the possibility of using this information during surgery to speed up treatment, escalate/deescalate surgery in specific phenotypic groups to introduce precision medicine to Neurosurgery.

Table of Contents

ACKNOWLEDGEMENTS	6
ABSTRACT	7
TABLE OF CONTENTS	8
LIST OF FIGURES	13
LIST OF TABLES	18
LIST OF PHOTOGRAPHS	20
LIST OF ABBREVIATIONS	21
1. BACKGROUND	22
1.1. CHAPTER OVERVIEW	22
1.2. PERSONAL BACKGROUND	22
1.3. THE HISTORICAL DEVELOPMENT OF BRAIN TUMOUR SURGERY	23
1.3.1. <i>Pioneering surgery</i>	23
1.3.2. <i>The birth of neurosurgery</i>	25
1.3.3. <i>Imaging technology</i>	26
1.4. CONCLUSION	28
2. INTRODUCTION	29
2.1. OVERVIEW	29
2.2. THE MOLECULAR BASIS OF BRAIN TUMOURS	29
2.2.1. <i>Epidemiology and aetiology</i>	29
2.2.2. <i>Clones and behaviour</i>	33
2.3. NEURO-ONCOLOGICAL SURGERY: CURRENT PRACTICE	34
2.3.1. <i>Functional-oncological balance</i>	35
2.3.2. <i>Tissue characterisation</i>	37
2.3.3. <i>Guidelines</i>	38
2.3.4. <i>Current practice</i>	40
2.3.5. <i>Summary</i>	41
2.4. PRECISION MEDICINE AND MACHINE LEARNING IN NEUROSURGERY	41
2.4.1. <i>Definition</i>	42
2.4.2. <i>Prerequisites</i>	43
2.4.3. <i>Promise for gliomas</i>	44
2.4.4. <i>New models</i>	45
2.4.5. <i>Summary</i>	46
2.5. iKNIFE: CLINICAL MASS SPECTROMETRY AND PRECISION MEDICINE	47
2.5.1. <i>Introduction</i>	47
2.5.2. <i>Technology</i>	48
2.5.2.1. Basic principles of mass spectrometry	48
2.5.2.2. Non-ambient and ambient mass spectrometry	48
2.5.2.3. Development of REIMS (the iKnife)	49
2.5.2.4. Model generation: the concept of profiling	50
2.5.3. <i>Clinical Applications</i>	50
2.5.3.1. Oncological surgery: introduction	50
2.5.3.2. Breast cancer	51
2.5.3.3. Brain tumours	53
2.5.3.4. Gastroenterology	56
2.5.3.5. Microbiology	57
2.5.4. <i>Conclusions</i>	59
2.6. RAMAN SPECTROSCOPY: AN INTRODUCTION	60

2.6.1. <i>Innovation after innovation</i>	60
2.6.2. <i>The chemistry of light scattering</i>	61
2.6.3. <i>Modern Raman systems and applications</i>	62
2.6.4. <i>Raman Spectroscopy in biological systems</i>	63
2.6.5. <i>The need for new Raman technology</i>	63
2.7. THE HYPOTHESIS	64
2.8. THESIS OBJECTIVES AND OUTLINE	64
3. SYSTEMATIC REVIEW ON INTRAOPERATIVE TISSUE DETECTION	66
3.1.1. <i>Introduction</i>	66
3.1.2. <i>Methods</i>	66
3.1.2.1. Search string	66
3.1.2.2. Inclusion criteria	67
3.1.2.3. Exclusion criteria	67
3.1.2.4. Study quality assessment	67
3.1.2.5. Data collection	68
3.1.3. <i>Results</i>	68
3.1.4. <i>Search strategy and selection criteria</i>	68
3.1.5. <i>Study demographics</i>	69
3.1.6. <i>Main results</i>	71
3.1.7. <i>Conclusions</i>	73
4. RAMAN SPECTROSCOPY	78
4.1. OVERVIEW	78
4.2. METHOD DEVELOPMENT	79
4.2.1. <i>Introduction</i>	79
4.2.2. <i>Method Development</i>	79
4.2.2.1. Laser wavelength	79
4.2.2.2. Sterilisation	80
4.2.2.3. Spectral range	80
4.2.2.4. Background fluorescence	80
4.2.2.5. Interference at the operative site	81
4.2.2.6. Light Interference from instrument displays	82
4.2.2.7. Ambient background residual	82
4.2.2.8. Laser power and distance from tissue	85
4.2.2.9. Blood artefact	87
4.2.2.10. Guidelines for good quality spectra	88
4.2.3. <i>Conclusion</i>	89
4.3. SYSTEM CALIBRATION	90
4.3.1. <i>Introduction</i>	90
4.3.2. <i>Calibration</i>	90
4.3.2.1. Wavelength calibration	90
4.3.2.2. Intensity calibration	91
4.3.2.3. Calibration checks	92
4.3.2.4. Interpretation of non-intensity calibrated spectra	93
4.3.3. <i>Summary</i>	96
4.4. SYSTEM SETUP	97
4.4.1. <i>Introduction</i>	97
4.4.2. <i>System overview</i>	97
4.4.3. <i>Reusable flexible Raman probe</i>	97
4.4.4. <i>Coupling of probe to laser and spectrometer</i>	98
4.4.5. <i>Optical filtering of the tissue emission</i>	98
4.4.6. <i>Optical filtering of the tissue excitation</i>	99
4.4.7. <i>System control</i>	100
4.4.8. <i>System packaging</i>	100
4.4.9. <i>Summary</i>	101
4.5. RAMAN SURGICAL PROTOCOL	102
4.5.1. <i>Introduction</i>	102
4.5.2. <i>Study design</i>	102

4.5.3. <i>Methods</i>	104
4.5.4. <i>Matching spectra to MRI and tissue samples</i>	105
4.5.5. <i>Quality control</i>	106
4.5.6. <i>in vivo brain measurement</i>	106
4.5.7. <i>Histopathological analysis</i>	108
4.5.8. <i>Conclusion</i>	108
4.6. DATA ANALYSIS METHODOLOGY	109
4.6.1. <i>Introduction</i>	109
4.6.2. <i>General principles</i>	109
4.6.2.1. Signal pre-processing	111
4.6.2.2. Dimensionality reduction	112
4.6.2.3. Classification using machine learning	113
4.6.2.4. Cross validation	115
4.6.3. <i>Dimensionality Reduction: PCA and LDA</i>	115
4.6.4. <i>Support Vector Machine Learning</i>	118
4.6.5. <i>Ensemble methods</i>	120
4.6.6. <i>PLS-DA</i>	121
4.6.7. <i>Classifier Performance Evaluation</i>	121
4.6.8. <i>Feature Extraction/ Peak Analysis</i>	123
4.6.9. <i>Summary</i>	123
4.7. RAMAN SPECTROSCOPY RESULTS	124
4.7.1. <i>Introduction</i>	124
4.7.2. <i>Demographics</i>	124
4.7.2.1. Introduction	124
4.7.2.2. Demographic analysis	124
4.7.2.3. Raman spectroscopy summary: spectral data	128
4.7.2.4. Summary	129
4.7.3. <i>in vivo Raman Spectroscopy of the Brain</i>	130
4.7.3.1. Introduction	130
4.7.3.2. Spectral Processing and Features	130
4.7.3.3. Summary	132
4.7.4. <i>in vivo vs ex vivo</i>	133
4.7.4.1. Introduction	133
4.7.4.2. Multivariate analysis	133
4.7.4.3. Summary	135
4.7.5. <i>Tumour Vs Normal</i>	135
4.7.5.1. Introduction	135
4.7.5.2. Average spectra	136
4.7.5.3. Multivariate analysis	138
4.7.5.4. Summary	160
4.7.6. <i>Tumour Density</i>	161
4.7.6.1. Introduction	161
4.7.6.2. Multivariate analysis	161
4.7.6.3. Summary	164
4.7.7. <i>Tumour Grade</i>	165
4.7.7.1. Introduction	165
4.7.7.2. Average spectra	165
4.7.7.3. Multivariate analysis	166
4.7.7.4. Summary	171
4.7.8. <i>Tumour Type</i>	171
4.7.8.1. Introduction	171
4.7.8.2. Average spectra	172
4.7.8.3. Multivariate analysis	173
4.7.8.4. Summary	177
4.7.9. <i>Tumour Margin</i>	177
4.7.9.1. Introduction	177
4.7.9.2. Data analysis	177
4.7.9.3. Summary	180
4.7.10. <i>Genomic Status</i>	181
4.7.10.1. Introduction	181

4.7.10.2. IDH	182
4.7.10.3. ATRX	191
4.7.10.4. MGMT	199
4.7.10.5. Summary	208
4.7.11. Conclusion	209
5. RAPID EVAPORATIVE IONIZATION MASS SPECTROMETRY (REIMS)	210
5.1. CHAPTER OVERVIEW	210
5.2. METHOD DEVELOPMENT	211
5.2.1. Introduction	211
5.2.2. Instrument location	211
5.2.3. Tubing characteristics	213
5.2.4. Bipolar type	214
5.2.5. IPA vs no IPA	215
5.2.6. Effect of Blood on the spectrum	217
5.2.7. Conclusions	218
5.3. CLINICAL TRIAL PROTOCOL	219
5.3.1. Introduction	219
5.3.2. Instrument setup, Calibration and Troubleshooting	219
5.3.3. Collecting in vivo data: Surgical Sampling	219
5.3.4. Pathology Specimen Processing	221
5.3.5. Summary	222
5.4. DATA ANALYSIS	223
5.4.1. Introduction	223
5.4.2. Data sources and integration	223
5.4.3. Data pre-treatment	223
5.4.4. Model building	224
5.4.5. Model performance	225
5.4.6. Conclusions	225
5.5. RESULTS	226
5.5.1. Introduction	226
5.5.2. Demographics	226
5.5.2.1. Introduction	226
5.5.2.2. Demographic analysis	227
5.5.2.3. Summary	231
5.5.3. In vivo vs Ex vivo	231
5.5.3.1. Introduction	231
5.5.3.2. Multivariate analysis	231
5.5.3.3. Conclusion	232
5.5.4. IPA vs Non-IPA	232
5.5.4.1. Introduction	232
5.5.4.2. Multivariate analysis	232
5.5.4.3. Conclusion	233
5.5.5. Tumour Vs Normal	234
5.5.5.1. Introduction	234
5.5.5.2. Average analysis	234
5.5.5.3. Multivariate analysis	235
5.5.5.4. Conclusion	240
5.5.6. Tumour Density	240
5.5.6.1. Introduction	240
5.5.6.2. Multivariate analysis	241
5.5.6.3. Conclusion	243
5.5.7. Tumour Type	244
5.5.7.1. Introduction	244
5.5.7.2. Average analysis	244
5.5.7.3. Multivariate analysis	245
5.5.7.4. Conclusion	249
5.5.8. Tumour Grade	249
5.5.8.1. Introduction	249

5.5.8.2. Multivariate analysis	249
5.5.8.3. Conclusion	255
5.5.9. <i>Genomic Studies</i>	256
5.5.9.1. Introduction	256
5.5.9.2. IDH	256
5.5.9.3. ATRX	260
5.5.9.4. MGMT	261
5.5.9.5. Summary	262
5.5.10. <i>Conclusion</i>	262
6. DISCUSSION	263
6.1. INTRODUCTION	263
6.2. CONTRIBUTION TO THE FIELD OF NEUROSURGICAL ONCOLOGY	263
6.2.1. <i>Development of 2 new modalities for surgery</i>	264
6.2.1.1. Raman	264
6.2.1.2. iKnife	265
6.2.2. <i>Concept and methodology for real time molecular profiling</i>	266
6.2.2.1. Data source	267
6.2.2.2. Data collection	267
6.2.2.3. Model generation	267
6.2.2.4. Model testing	268
6.2.3. <i>A new platform for precision medicine</i>	268
6.3. IN VIVO SPECTROSCOPY FOR THE HUMAN BRAIN	269
6.3.1. <i>Raman spectroscopy</i>	269
6.3.1.1. Tumour detection	274
6.3.1.2. Tumour density	277
6.3.1.3. Tumour grade	278
6.3.1.4. Tumour type	278
6.3.1.5. Tumour margin	281
6.3.1.6. Genetic status	282
6.3.2. <i>iKnife: REIMS</i>	282
6.3.2.1. Tumour detection	283
6.3.2.2. Tumour density	283
6.3.2.3. Tumour type	284
6.3.2.4. Genetic status	284
6.3.3. <i>Conclusion</i>	285
6.4. LIMITATIONS OF THIS THESIS	286
6.5. FUTURE TRIALS AND DEVELOPMENTS	288
6.5.1. <i>Future development of Raman</i>	288
6.5.1.1. Hardware	288
6.5.1.2. Software	289
6.5.2. <i>Future development of iKnife</i>	290
6.5.3. <i>UK multicentre feasibility study</i>	290
6.5.4. <i>Conclusions</i>	292
7. REFERENCES	ERROR! BOOKMARK NOT DEFINED.
8. APPENDIX	301
8.1. APPENDIX 1: RAMAN IN VIVO DATA ENTRY SHEET	301
8.2. APPENDIX 2: IN VIVO SPECIMEN LABELLING GUIDE	302
8.3. APPENDIX 3: IKNIFE SETUP, CALIBRATION AND TROUBLESHOOTING	303
8.3.1. <i>Instrument calibration</i>	306

List of Figures

Aim	7
Rationale.....	7
Methods	7
Results	7
Conclusions.....	7
Figure 1. Flow chart of the new diagnoses integrating histological and molecular information in the new WHO classification of CNS tumours.....	31
Figure 2. Diagnosis and treatment guideline pathway from the European Association for Neuro-Oncology (EANO).....	32
Figure 3. Integrating the molecular evolution of glioma throughout the treatment pathway. (Aldape et al., 2018).	34
Figure 4. Goals of Neuro-Oncological Surgery.	36
Figure 5. Data flow in molecular profiling of human tumour tissue.....	50
Figure 6. ex vivo and intraoperative workflow: breast cancer iKnife (St John et al., 2017b)	52
Figure 7. REIMS diagram in the operating room, (Balog et al., 2013)	55
Figure 8. LDA analysis of data from iKnife in a range of tumours. (Balog et al., 2013).....	55
Figure 9. in vivo endoscopy using the iKnife.	57
Figure 10. A timeline of the key discoveries leading to present day uses of Raman spectroscopy.....	61
Figure 11. Raman spectrometer schematic.	62
Figure 12. PRISMA 2009 flow diagram showing the study selection process for meta-analysis.....	69
Study demographics.....	69
Figure 13. Bar chart showing number of <i>in vivo</i> and <i>ex vivo</i> studies used for intraoperative brain tissue detection.....	70
Figure 14. Distribution of studies with respect to technology category for intraoperative brain tissue detection.	71
Mass spectrometry.....	73
Figure 15. Fluorescence from artificial sources.....	83
Figure 16. Raman spectrum obtained in the operating theatre on brain tissue with significant fluorescent light background from mixed IR sources.	83
Figure 17. Raman spectrum of the human brain taken during surgery with IR precautions.	85
Figure 18. Raman spectra from patient 1.	86
Figure 19. Example of Raman spectrum with blood artefact.....	87
Figure 20. Raman spectrum taken after haemostasis.	88
Figure 21. Wavelength calibration of the Raman system.	91
Figure 22. Intensity calibration of the Raman system.	92
Figure 23. Hypothetical flat emission profiles at 1, 2, 3, 4, 5, and 6 x 10 ⁴ counts measured by the Raman system.	94
Figure 24. Residuals after subtracting the intensities derived from iterative polynomial fit to the measured emissions of hypothetical flat emission profiles at 1, 2, 3, 4, 5, and 6 x 10 ⁴ counts.	95
Figure 25. Average tumour, normal and difference spectra using data from all patients.	96
Figure 26. Schematic diagram of the Raman system (not to scale).....	99
Figure 27. Raman study pathway diagram.....	103
Figure 28. Flow chart of ex vivo sampling in the operating theatre.	104
Figure 29. Data flow paradigm in spectroscopy, taken from Baker et al. (Baker et al., 2018).....	110
Figure 30. Machine learning methods grouped into methodological classes.	114
Figure 31. PCA plot using gscatter showing breakdown by tumour type.	117
Figure 32. PCA plot using gscatter showing breakdown by tumour grade and type.	118
Figure 33. Classification using SVM: 2 classes separated by a linear margin.	119
Figure 34. A 2x2 confusion matrix of a binary classifier.....	122
Figure 35. An example of a ROC curve calculated for tumour vs non-tumour classification using Raman data.	123

Figure 36. Histogram representing age distribution of recruited patients.	126
Figure 37. Number of patients recruited for each gender.	126
Figure 38. Distribution of histology findings amongst recruited patients.	127
Figure 39. Distribution of WHO grades (2016 classification) amongst recruited patients.	127
Figure 40. Demographics of the spectra collected in the study.	128
Figure 41. Example of a two range (394 – 670 cm ⁻¹ and 672 – 1870 cm ⁻¹) iterative polynomial fit to a calibrated and smoothed tissue emission spectrum.	131
Figure 42. Raman spectrum of <i>in vivo</i> brain tissue of different WHO grades.	132
Figure 43. Characteristics of spectra collected from <i>in vivo</i> and <i>ex vivo</i> sources.	133
Figure 44. LDA plot of <i>in vivo</i> and <i>ex vivo</i> data.	134
Figure 45. Average normal vs average tumour spectrum.	136
Figure 46. Comparison of normalized <i>in vivo</i> Raman spectra from brain tissue and Raman spectra from biochemical standards expected to be found in brain tissue in abundant quantities.	137
Figure 47. PCA of all 1171 spectra from brain tissue obtained during surgery.	138
Figure 48. PCA of Raman spectra with biopsy information only.	139
Figure 49. PCA of tumour vs normal <i>in vivo</i> spectra, excluding GBM.	140
Figure 50. Posterior probability plots obtained from the analyses of 1171 Raman spectra using PLS with waveband selection.	140
Figure 51. Boxplot representations of the posterior probability distributions as determined by PLS with waveband selection analyses of 1171 Raman spectra.	141
Figure 52. The ROC curve (continuous curve) and 95% CIs (dashed curves) derived from the posterior probabilities determined by PLS with waveband selection analyses of 1171 Raman spectra.	141
Figure 53. Average auto-fluorescence spectra from brain tissue derived from polynomial fits to the calibrated raw spectra.	143
Figure 54. Average normalized auto-fluorescence spectra from brain tissue derived from polynomial fits to the calibrated raw spectra.	144
Figure 55. Posterior probability plots obtained from the analyses of 1171 auto-fluorescence spectra using PLS with waveband selection.	145
Figure 56. The ROC curve (continuous curve) and 95% CIs (dashed curves) derived from the posterior probabilities determined by PLS with waveband selection analyses of 1171 auto-fluorescence spectra.	145
Figure 57. Boxplot representations of the posterior probability distributions as determined by PLS with waveband selection analyses of 1171 auto-fluorescence spectra.	146
Figure 58. Posterior probability plots obtained from the analyses of 1171 combined Raman and auto-fluorescence spectra using PLS with waveband selection.	147
Figure 59. Boxplot representations of the posterior probability distributions as determined by PLS with waveband selection analyses of 1171 combined Raman and auto-fluorescence spectra.	148
Figure 60. The ROC curve (continuous curve) and 95% CIs (dashed curves) derived from the posterior probabilities determined by PLS with waveband selection analyses of 1171 combined Raman and auto-fluorescence spectra.	148
Figure 61. Posterior probability plots from 953 combined Raman and auto-fluorescence spectra taken from gliomas.	150
Figure 62. Box plot representations of the posterior probability distributions as determined by PLS with waveband selection from 953 combined Raman and auto-fluorescence spectra from gliomas.	150
Figure 63. The ROC curve (continuous curve) and 95% CIs (dashed curves) derived from the posterior probabilities determined by PLS with waveband selection analyses of 953 combined Raman and auto-fluorescence spectra taken from patients with gliomas.	151
Figure 64. ROC curve of the Random Forrest model classifying tumour vs normal spectra from all 1171 spectra.	154
Rapidminer	157
Rattle	158
Figure 65. Confusion matrix and ROC curve of the cubic SVM classifier for 2226 <i>in vivo</i> and <i>ex vivo</i> spectra.	160
Figure 66. PCA plot of Raman spectra from biopsied tissue labelled for tissue density.	162
Figure 67. PCA plot of Raman spectra with fluorescence from biopsied tissue labelled for tissue density.	163
Figure 68. LDA plots of Raman spectra (left) and Raman and fluorescence spectra (right) for tumour density.	163
Figure 69. Average Raman spectra of the different tissue types included in the Raman study.	166
Figure 70. Average Raman spectra of WHO grade 2 and grade 3 tumours compared.	166

Figure 71. PCA plot of the first 3 principle components determining variation of WHO grades in biopsy-only spectra.	167
Figure 72. LDA plot of WHO grades in all Raman spectra collected in the study.	167
PLS with waveband selection predicted the pathology of these cases with a sensitivity of 85% (95% CI of 82-88%), and a specificity of 80% (Table 14). See Table 15 for a breakdown of results for different tumour subclasses that I obtained.....	167
Figure 73. ROC curve of WHO grade as predicted by SVM models on <i>in vivo</i> and <i>ex vivo</i> data combined.....	171
Figure 74. Averaged glioma spectra compared with non-glioma spectra from 29 patients.	172
Figure 75. PCA of all in vivo and ex vivo spectra.	173
Figure 76. LDA plot of glioma vs non-glioma for all in vivo and ex vivo spectra combined.	174
Figure 77. PCA plot showing PCA components 3,4 and 5.	175
Figure 78. LDA plot of tumour type.	175
Figure 79. Predicted classes from SVM model for P5 data.	178
Figure 80. Predicted classes from Ada Boost model for P5 data.....	179
Figure 81. 3D visualisation of a brain tumour showing Raman data analysis points and highlighting the radiological tumour border, A/P view.	179
Figure 82. 3D visualisation of a brain tumour showing Raman data analysis points and highlighting the radiological tumour border, lateral view.	180
Figure 83. Average spectra of IDH mutated and non mutated tissue (left) and the difference spectrum (right).	182
Figure 84. IDH spectral comparisons and associated sensitivity and specificities for spectra grouped by grade.	183
Figure 85. Comparison of successive difference spectra for IDH mutation with reference spectra.	184
Figure 86. The effect of the IDH MT is investigated in spectra from all tumours but again only those with ATRX 1, and MGMT 0.	185
Figure 87. Second order derivative peaks and comparison with 2-HG reference spectrum.....	185
Figure 88. PCA plot of IDH MT vs IDH WT.	186
Figure 89. Loadings plot of the PCA analysis in Figure 88.....	186
Figure 90. Precision/Recall plot comparing SVM and ensemble methods model performance.	189
Figure 91. Probability of IDH mutation in brain tissue using PCA-LDA prediction models.....	189
Figure 92. Frequency distribution in the probability of brain tissue having the IDH mutation using PCA-LDA..	190
Figure 93. Distribution of number of spectra within probability width of 0.1 against posterior probability of spectra acquired from IDH mutated tissue.....	190
Figure 94. ROC curve and AUC for the PLS-LDA model when predicting IDH mutation.	191
Figure 95. Average spectra from ATRX WT and ATRX MT tissue.....	192
Figure 96. ATRX Raman difference spectrum compared to DNA reference spectrum.....	192
Figure 97. Supervised hierarchical cluster analysis of ATRX spectra.	193
Figure 98. Comparison of successive difference spectra for ATRX mutation with reference spectra.	194
Figure 99. PCA plot of ATRX MT vs ATRX WT.	194
Figure 100. Loadings plot of the PCA analysis for ATRX mutation.....	195
Figure 101. Sensitivity/Specificity plot comparing SVM and ensemble methods model performance for ATRX status.....	197
Figure 102. Probability of ATRX mutation in brain tissue using PCA-LDA prediction models.	198
Figure 103. Frequency distribution in the probability of brain tissue having the ATRX mutation using PCA-LDA.	198
Figure 104. ROC curve and AUC for the PLS-LDA model when predicting ATRX mutation.	199
Figure 105. Average spectra from MGMT methylated and MGMT unmethylated tissue.	199
Figure 106. MGMT Raman difference spectrum.....	200
Figure 107. Supervised hierarchical cluster analysis of MGMT spectra.....	201
Figure 107b. MGMT Raman difference spectrum compared to DNA reference spectrum.	201
Figure 108. Comparison of successive difference spectra for MGMT methylation status with reference spectra.	202
Figure 109. PCA plot of Raman spectra from MGMT methylated vs non-methylated tissue.	203
Figure 110. Sensitivity/Specificity plot comparing SVM and ensemble methods model performance for MGMT status.....	205
Figure 111. ROC curve and AUC for the PLS-LDA model when predicting IDH mutation.	206
Figure 112. Frequency distribution in the probability of brain tissue having the MGMT methylation using PCA-	

LDA	207
Figure 113. Probability of MGMT mutation in brain tissue using PCA-LDA prediction models.....	207
Figure 114. ROC curve and AUC for the PLS-LDA model when predicting MGMT status.	208
Figure 115. Mass spectra of ERBE versus Hygitech tubing.....	213
Figure 116. Above: Mass spectra obtained using Braun forceps. Below: Mass spectra obtained using Erbe forceps.....	214
Figure 117. Erbe Bipolar irrigation forceps.	215
Figure 118. Calf brain REIMS spectra illustrating the effect of IPA.....	215
Figure 119. ex vivo testing of IPA using brain tumour tissue from patient 37.	216
Figure 120. ex vivo testing of alcohol solvents on pork tissue.	217
Figure 121. Mass spectrum obtained when sampling fresh haematoma forming during surgery.....	218
Figure 122. The use if the iKnife during brain surgery.....	220
Figure 123. Progressive sampling during an iKnife case.	220
Figure 124. Demographic summary of all patients in the iKnife trial	227
Figure 125. Detailed breakdown of the demographic characteristics of the spectra collected.....	228
Figure 126. PCA plot of the top 3 principle components comparing Gender of patients 9n the study.....	229
Figure 127. LDA plot of the top 3 LDA components comparing Gender of patients in the study.	229
Figure 128. PCA plot of the top 3 principle components comparing the age of patients in the study.....	230
Figure 129. LDA plot of the top 3 LDA components comparing the age of patients in the study.	230
Figure 130. 3-D representation of PCA of <i>in vivo</i> vs <i>ex vivo</i> iKnife spectra. LDA analysis of <i>in vivo</i> vs <i>ex vivo</i> datasets. There is no clear separation between both groups.....	232
Figure 131. PCA analysis of IPA vs non-IPA spectra.....	233
Average Spectra	234
Figure 132. Averaged mass spectra for normal and tumour tissue.....	235
For multivariate analysis the spectra were grouped into non-IPA spectra and IPA spectra. Non-IPA results will be presented first.....	235
Figure 133. PCA analysis of tumour vs normal samples, non-IPA, m/z100-1000.....	236
Figure 134. LDA analysis of tumour vs normal samples, non-IPA, m/z100-1000.....	236
Figure 135. Non-IPA PCA-LDA model loadings for normal vs tumour.	238
Figure 136. LDA plot of tumour vs normal for IPA spectra, m/z 100-1000.....	238
Score and Loading Plots.....	239
Figure 137. Scores and loadings from the combined IPA-nonIPA PCA-LDA model.	239
Figure 138. PCA plot of tumour cell density, none vs high, non-IPA spectra, m/z 100-1000.....	241
Figure 139. PCA plot of moderate vs high density tumour tissue spectra. IPA, m/z100-1000.....	242
Figure 140. LDA plot of tumour density using IPA and non-IPA spectra, m/z 600-800.	243
Figure 141. Average mass spectra, m/z range 100-1000, for glioma (top) and non-glioma tissue (bottom), non-IPA.	244
Figure 142. Comparison of the average spectrum of common brain tumours using REIMS, m/z 500-1000.	245
Figure 143. PCA analysis of glioma vs non-glioma in 16 selected patients matched for age and gender, m/z600-1000, non-IPA.....	246
Figure 144. LDA analysis of glioma vs non-glioma in 16 selected patients matched for age and gender, m/z600-1000, non-IPA.....	246
Figure 145. Loadings plot of the PCA-LDA model for non-IPA glioma vs non-glioma REIMS model.....	247
Figure 147. PCA analysis of tumour type histology, m/z600-1000, non-IPA.....	248
Figure 148. LDA analysis of tumour type histology, m/z600-1000, non-IPA.....	248
Cross validation results of the PCA-LDA model in Table 31 reveals a low accuracy of 47.1% with leave 20% out cross validation.	248
Model LOPOCV Leave 20% Out CV	248
Figure 149. PCA plot of low grade and high grade REIMS spectra. IPA, m/z 600-800	250
Figure 150. LDA plot of WHO grades from non-IPA and IPA REIMS spectra.	250
Figure 151. LDA plot of WHO grades from IPA REIMS spectra.....	251
Figure 152. PCA plot using 300 components of tumour grade, REIMS IPA and non-IPA combined.....	251
Figure 153. LDA plot using 300 components of tumour grade, REIMS IPA and non-IPA combined.....	252
Figure 154. LDA plot of tumour grade for IPA results only, m/z range 100-1000.	253
Figure 155. Tumour grade PCA-LDA loadings for the IPA model.....	253
Figure 156. LDA plot of IDH mutation in GBM, non-IPA, m/z100-1000.....	257
Figure 157. LDA plot of IPA data and IDH mutation status using REIMS.	258

Figure 158. The loadings plot for the IDH mutation model using REIMS data.	259
<i>LOPOCV (top) and leave 20% cross validation (bottom) was performed on the PCA-LDA model.</i>	261
Figure 159. Averaged Raman spectra from the study by Jermyn et al. (Jermyn et al., 2015a)	270
Figure 42. Raman spectrum of in vivo brain tissue of different WHO grades.	271
Figure 160. Photomicrographs (H&E staining) of healthy (A) and glioma (B–D) brain tissue sections.....	272
Figure 161. The mean <i>ex vivo</i> Raman spectra of normal tissue, medulloblastoma and glioma tissue (top) and mean spectra for each type of tumour (bottom).	273
Figure 162. Significant Raman spectral characteristics for brain tumours- reproduced from Leslie et al. (Leslie et al., 2012)	274
Figure 163. Raman spectra of normal brain tissue and necrotic tissue.	274
Figure 53. Average auto-fluorescence spectra from brain tissue derived from polynomial fits to the calibrated raw spectra.	276
Figure 65. ROC curve of the cubic SVM classifier for 2226 <i>in vivo</i> and <i>ex vivo</i> spectra.	277
Figure 164. Average <i>in vivo</i> Raman spectra of GBM and astrocytoma tissue.	279
Figure 165. Average <i>ex vivo</i> GBM and normal spectra from Kalkanis et al. (Kalkanis et al., 2014)	279
Figure 166. Normal and tumour spectra from single patients from Jermyn et al.	280
Figure 77. PCA plot showing PCA components 3,4 and 5.	280
Figure 81. 3D visualisation of a brain tumour showing Raman data analysis points and highlighting the radiological tumour border , A/P view.	281
Figure 167. LDA plot of <i>ex vivo</i> samples from the first use of the iKnife (Balog et al., 2013).....	283
Figure 168. Background signal on REIMS machine.....	304
Figure 169. Calibration screen on the iKnife.....	307
Figure 170. Calibration check pass screen for the iKnife.	308

List of Tables

Table 1. Simple averages of sensitivity and specificity for IMA techniques in breast cancer surgery. (St John et al., 2017a)	52
Table 2. MS/MS based phospholipid identification in breast cancer. (St John et al., 2017b)	53
Table 3. The broad groups of technology and subtypes of technology used for intraoperative brain tissue detection.....	70
Table 4. Diagnostic accuracy of intraoperative brain tumour tissue detection.	73
Table 5. Summary of technologies used for intraoperative brain tumour detection.	74
Table 6. Comparing machine learning methods for chemometrics in Neurosurgery. Taken from Senders et al. (Senders et al., 2018).....	111
Table 7 Summary table of all patients recruited to the Raman spectroscopy study.	125
Table 8. Distribution of numbers of <i>in vivo</i> spectra amongst different tumour grades and diagnosis types. ...	129
Table 9: Classification of spectra acquired from patients with different tumour subclasses based on different multivariate methods.	142
Table 10. Classification of different tumour subclasses based on PLS analyses with waveband section.	149
Table 12. Classification of different tumour subclasses based on PLS analyses with waveband section.	152
Table 13. Classification of spectra acquired from patients with different tumour subclasses based on different multivariate methods.	152
Table 14. Classification of spectra acquired from patients with different tumour subclasses based on different multivariate methods.	168
Table 15. Classification of different tumour grade based on PLS analyses with waveband section.....	169
Table 16. Classification of different tumour grade based on PLS analyses with waveband section.....	176
Table 17. Patient demographics and histological diagnosis for genetic analysis.	181
Table 18. Summary of IDH, ATRX and MGMT mutational states and the number of spectra collected from tissue for molecular analysis.....	182
Table 19. Key peaks seen in principle components 3,4 and 18.	187
Table 20. Key peaks seen in principle components 2,5 and 7.	195
Table 20a. Key peaks seen in principle components 2,3 and 4.	203
Table 21. Partial Least Squares/Linear Discriminant Analysis (PLS-LDA) performance using Raman spectroscopy for the prediction of key genetic and epigenetic changes in glioma.	209
Table 22. Confusion matrix from cross validation of IPA vs non-IPA using a PCA model, m/z 100-1000.	233
Table 23. Confusion matrix of the LOPOCV results for IPA vs Non-IPA comparison, PCA-LDA model, m/z 100-1000.....	237
Table 24. Confusion matrix of the LOPOCV results for the non-IPA PCA/LDA model, m/z 600-1000.....	237
Table 25. Confusion matrix of leave 20% out cross validation of the PCA-LDA model of IPA data for tumour vs normal.	239
Table 26. LOPOCV confusion matrix of tumour vs normal tissue, IPA and non-IPA data combined, m/z 100-1000, lockmass 554.2615.....	240
Table 27. Leave 20% out cross validation of the tumour density non-IPA model.	241
Table 28. Confusion matrix from LOPOCV of the moderate vs high tumour density IPA model.....	242
Table 29. LOPOCV results for non-IPA REIMS data.....	247
Table 30. LOPOCV results for IPA REIMS data.....	247
Table 31. Cross validation performance of PCA-LDA models using IPA, non IPA and combined data.....	249
Table 32. LOPOCV of tumour grade low vs high PCA-LDA model, non-IPA, m/z 600-800.	254
Table 33. LOPOCV of tumour grade PCA-LDA model, non-IPA, m/z 600-800.	255
Table 33. Leave 20% out cross validation of the IDH mutation REIMS model.	257
Table 35. Leave 20% out cross validation on the PCA-LDA model for IDH mutation using IPA data.....	258
Table 36. Leave one patient out cross validation results of the combined IPA and non-IPA REIMS model for IDH	

mutation.	259
Table 37. Leave 20% out cross validation (top) and leave one patient out cross validation (bottom) for IDH mutation in GBM patients using REIMS data.	260
Table 38. ATRX cross validation using the PCA-LDA model for REIMS tumour data.	260
Table 39. Cross validation results of ATRX mutation from REIMS data using a matched training set.	261
Table 40. Results of LOPOCV for the MGMT methylation REIMS PCA-LDA model.	261
Table 41. LOPOCV for MGMT methylation using the modified REIMS PCA-LDA model.	262
Table 42. SVM classifications for DESI on glioma samples for a single patient from Eberlin et al. (Eberlin et al., 2013)	284
Table 43. Summary table comparing diagnostic performance and system features of Raman spectroscopy versus iKnife.	285

List of Photographs

Photograph 1. Customized TECAN Freedom EVO automated sample handling machine and sampling probe. ...	59
Photograph 2. Theatre 7 setup highlighting sources of infrared.....	81
Photograph 3. Photograph showing the complete system in use mounted on a mobile cart.	101
Photograph 4. The layout of Theatre 7, Charing Cross Hospital.....	212

List of Abbreviations

Abbreviation	Description
5-ALA	5-Aminolevulinic Acid
ATRX	Alpha-Thalassemia/Mental Retardation Syndrome, X-Linked
AUC	Area Under Curve
BOLD	Blood Oxygenation Level Dependent
BT	Boosted Trees
CBTRUS	Central Brain Tumor Registry of the United States
CT	Computed Tomography
CV	Cross Validation
DTI	Diffusion Tensor Imaging
EEG	Electroencephalography
EOR	Extent Of Resection
EPI	Echo-planar imaging
GBM	Glioblastoma Multiforme
IDH	Isocitrate Dehydrogenase
IPA	Isopropyl Alcohol
LDA	Linear Discriminant Analysis
LOOCV	Leave One Out Cross Validation
LOPOCV	Leave One Patient Out Cross Validation
MGMT	Methylguanine Methyltransferase
MRI	Magnetic Resonance Imaging
MRS	Magnetic Resonance Spectroscopy
PCA	Principle Component Analysis
PET	Positron Emission Tomography
PLS-DA	Partial Least Squares Discriminant Analysis
REIMS	Rapid Evaporative Ionization Mass Spectrometry
RF	Random Forrest
ROC	Receiver Operator Characteristic
SVM	Support Vector Machines
WHO	World Health Organisation

1. Background

1.1. Chapter Overview

The aim of this chapter is to establish the narrative which has driven this research. A personal account of my own training in brain tumour surgery will set the scene for a brief history of surgery for brain tumours. The current technologies used in surgery will be traced from their origins to current use in the operating theatre. The limitations of current technologies will be discussed.

1.2. Personal background

At the time of undertaking this thesis I am a senior trainee in Neurosurgery on the North Thames Neurosurgery Rotation. I graduated in 2003 after studying medicine at New College, University of Oxford where exposure to Sir William Osler's thinking heavily influenced my approach to clinical medicine. I entered my pre-registration house officer year and worked in the Oxford and South Yorkshire deaneries. I completed my Basic Surgical Training in Cambridge in 2006 and entered Neurosurgery speciality training in 2007 leading to the award of the Fellowship of the Royal College of Surgeons in Neurosurgery in 2014 and completion of training leading to the Certificate of Completion of Training in Neurosurgery and entry into the Specialist Register in November 2017.

During my 10 years of training in neurosurgery I became painfully aware of the constraints on the surgeon whilst operating on brain tumours. The task when operating on intrinsic brain tumour seemed almost impossible. The challenge is similar to that of removing a grape which is embedded in jelly. Except that the grape is the same colour as the jelly, with a texture very similar to that of jelly. Add to that the fact that the 'jelly' must be disturbed as little as possible as it contains the vital biological substrate that controls and coordinates speech and language, vision, movement of the body, and consciousness and then the dilemma of the neurosurgeon becomes clearer. It seemed astonishing to me that advanced methods of tissue detection during surgery to aid the surgeon were lacking. My training had made it plain to me that when a solution doesn't exist it must be found so that patients can be served better. The task of improving surgical techniques must start and finish with the patient. It was obvious to me as a surgeon that in order to remove tumours more completely and safely for my patient systems needed to be available to assist the surgeon in quickly identifying normal tissue and cancerous tissue. These systems must be fast, easy to use and interpret and easily form part of the operating theatre workflow. I set about finding colleagues and experts who would be able to help me as a surgeon in this quest to improve the surgery I could offer. I was

fortunate to encounter a mentor unlike any other. In Kevin O'Neill I found a surgeon with immense natural talent, ability, uncompromising standards and vision to transform our speciality. In Sergei Kazarian I found an expert in Raman spectroscopy who believed in multidisciplinary collaboration moving Raman into clinical practice. In Zoltan Takats I found a pioneer of mass spectrometry willing to push innovation into the operating theatre.

My hope is that this thesis will be a first step towards bringing the science of chemical measurement known as chemometrics into the operating theatre to enable neurosurgeons to operate with a revolutionary level of precision. Through this it is my belief that surgery can be further refined and made safer. It is also my intention that this work strengthens the position of the surgeon as someone with privileged access to human organs and systems who is in a prime position to initiate, manage and carry out multidisciplinary clinical trials which can transform our understanding of cancer *in situ* like no one else. For neurosurgery as my discipline it is my aspiration that this work will herald a new era of *in vivo* neurosurgical spectroscopy.

1.3. The Historical Development of Brain Tumour Surgery

Surgery for brain tumours has undergone an evolution over the past 130 years from a dangerous and often fatal endeavour to a safe and effective procedure offered to thousands of patients every year. The story of how it has developed has determined the current philosophy of the surgical approach to the treatment of brain tumours. This section will cover some details of the first documented operation on an intrinsic tumour and will then trace the subsequent improvements in technology that have directly led to the evolution of brain tumour surgery culminating in the current array of technologies at the disposal of today's neurosurgeon.

1.3.1. Pioneering surgery

The first documented operation on a primary brain tumour was performed on the 25th of November 1884 at The Hospital for Epilepsy and Paralysis, Regents Park, London by Rickman J Godlee (Bennett, 1885). The operation was performed at the behest of the neurologist Dr Hughes Bennet. Operations had previously been performed for meningiomas and osteomas, but this was the first example of surgery for an intrinsic brain tumour guided by neurological examination and with no obvious external landmarks to guide surgery. In an operation observed by none other than Hughlings Jackson, David Ferrier (both prominent neurologists at the time) and Victor Horsley (the father of Neurosurgery in the UK) a trephine was fashioned in the exact spot where the tumour was believed to be on the basis of neurological history and examination alone.

The patient was a 25 year old Scottish farmer afflicted by painful daily focal motor seizures of his left mouth, tongue and arm and progressive worsening weakness down the left side of his body. Based on the emerging knowledge of localisation of cerebral function Bennet determined a lesion in the right precentral gyrus was

probably to blame for the patient's symptoms and surgery was considered as a last resort in the face of almost certain death, and Mr Godlee was called in to operate.

The trephination performed over the right precentral gyrus revealed a transparent lobulated solid tumour which bulged under pressure. 2 further trephines were placed adjacent to the first to aid removal of the tumour and the growth was "easily separated" from the brain substance using a narrow spatula. Finger dissection was used to remove the rest of the tumour resulting in a cavity "of a size into which a pigeon's egg could fit".

Subsequent histological examination revealed this to be a glioma. The patient awoke from chloroform anaesthesia no worse neurologically than how he had been prior to surgery. The operation was deemed a success. The patient's postoperative course was, however, to take a turn familiar to those who have operated on the human nervous system. Despite technically excellent surgery, disaster can strike at any time rendering the initial meticulous efforts potentially worthless. 3 weeks later the patient was affected by rigors and high fever and a drop in consciousness. The condition of the patient continued to worsen, and he died on the 23rd of December 1884. Post mortem examination revealed evidence of meningitis but no residual tumour. The cause of death was a postoperative infection.

This example in some ways typified brain surgery at the time - despite valiant efforts and technically correct surgery, the human nervous system can be unforgiving of any complications. News of the operation quickly spread amongst the medical community both in the United Kingdom and across the Atlantic. In the words of Bennet:

"We have nowhere been able to discover the recorded example of a case where a cerebral tumour was diagnosed by the symptoms observed, without visible or tangential external signs, and was in consequence operated on and successfully removed."

Both Bennet and Godlee concluded that the case "leads us to believe that there is an encouraging prospect for the future of cerebral medicine and surgery, and that as a tumour of the brain can be diagnosed with precision and successfully removed without immediate danger to life". The triad of anaesthesia, antisepsis and cerebral localisation seems to have opened up the possibility for safe surgery on brain tumours.

The precedent was set and brain surgery was on a course of advancement. Sir William Osler commended the operation, it was almost certainly touched upon when Victor Horsley delivered his lecture "On the Technique of Operations on the Central Nervous System" in Toronto in 1906. At this time Osler asked for "medico-chirurgical neurologists, properly trained in anatomical, physiological, clinical and surgical aspects of the subject". One of the founding fathers of modern Neurosurgery, Harvey Cushing, was Osler's contemporary and a close student. Osler also inspired his Oxford student Wilder Penfield who would go on to found the neurological institute at McGill University. Following Victor Horsley's untimely demise of heat stroke whilst serving for the British Army in Mesopotamia the light of advancement for brain tumour surgery shifted to North America.

1.3.2. The birth of neurosurgery

Harvey Cushing was working hard in the early 1900's to develop neurosurgery as a speciality at Johns Hopkins Hospital. When he started there the mortality from brain tumour surgery was an astonishing 30-50% (Greenblatt, 2003). By 1910 he had operated on 180 tumour cases and documented a much reduced mortality of 10-15%. Through his high case load and applying meticulous techniques and note taking Cushing played a central role in improved mortality and morbidity for brain tumour surgery. Careful consideration of patient positioning, cranial openings and control of bleeding as well as the record keeping required to identify key factors in determining surgical outcomes enabled safer surgery. During this time surgery for brain tumours emerged as a feasible and safe endeavour. Further advancement towards new techniques in brain surgery came from Canada in the form of one of Cushing's contemporaries: Wilder Penfield.

Penfield pioneered the use of awake craniotomy as a tool for operating on patients with epilepsy and tumours. The main advantage of awake craniotomy over one performed under general anaesthesia is the ongoing ability to evaluate the function of critical anatomical structures, enabling the surgeon to avoid damage to these critical areas. This was a technique useful for patients with tumours close to eloquent brain areas serving speech and movement. Through using electrical stimulation on the surface of the brain of an awake patient Penfield was able to produce, for the first time, detailed maps of the organisation of the somatic and motor cortex. Penfield and co investigators Edwin Boldrey and Theodore Rasmussen went on to publish in 1937 what is considered to be one of the most famous conceptual maps in neuroscience (Penfield and Boldrey, 1937). Through the direct cortical stimulation of the brain in patients with local anaesthesia they were the first to be able to accurately map and distinguish motor and sensory homunculi. The strikingly graphical findings have impressed neuroscience students and researchers ever since.

No case better illustrates his commitment to awake craniotomy than when he performed the procedure on his own sister who was suffering from a glioma of the right frontal lobe. In an effort to prolong life by removing the tumour but also to preserve function by testing the brain on her as an awake patient, the operation gave his sister 2 more valuable years of life.

Along with awake craniotomy, other techniques were investigated by Penfield in an effort to aid localisation of function and prevent damage to the brain whilst operating. Electrical tracings from brain activity were first reported in 1928 and the first use of intraoperative electroencephalography (EEG) was reported by Foerster and Alternberger in 1935. Penfield used electrocorticography for localisation and surgical treatment of epilepsy during awake craniotomies and developed the field further. Intraoperative neurophysiology monitoring has progressed to become an important method of continuously monitoring brain activity in the anaesthetised and awake surgical patient.

Through the demonstration of cerebral localisation, Cushing's meticulous technique, and new forms of intraoperative monitoring of the awake and asleep patient the knowledge and practice of neurosurgery was advancing quickly in the 20th Century. Further advances would be directly connected to technology as new methods of

visualising the human brain extended the surgeons abilities even further.

1.3.3. Imaging technology

Visualisation of brain tissue is of central importance to the surgeon. A surgeon can only operate on what he or she can see and this has a direct impact on patient safety. The field of microsurgery developed in the 1960s and directly made advanced neurosurgery a reality, giving surgeons the confidence and ability for the first time to manipulate brain tissue at the microscopic level.

The Operative Microscope

The surgical microscope was increasingly being used by allied surgical specialities such as otolaryngology in the 1950s and 60s which set the precedent for neurosurgeons. Microsurgical laboratories in the United States allowed budding surgeons such as Yasargil to learn the skills required to use the technology on animal models prior to first use in humans. Yasargil's group in Zurich was to further develop the operative microscope by adding a counter balance to make it easier to move in the operative field. Electromagnetic brakes were used at each hinge point allowing full mobility with maximum stability. Yasargil went on to make the operating microscope an integral part of modern neurosurgery. It is hard to imagine any brain tumour operation where a microscope is not used by the surgeon to visualise the intricate structures found in the human brain and enable a finer level of tissue manipulation and hazard avoidance than would be otherwise possible.

A key feature of hazard avoidance is preoperative planning: knowing where vital structures such as blood vessels or eloquent brain regions are likely to be is crucial when operating. To this end crucial advances in medical imaging technology were about to occur which propelled surgery for brain tumours into another age; that of image guided surgery.

Computed Tomography (CT) Scanning

CT scanning was invented and introduced into clinical practice by Sir Godfrey Hounsfield. On 1 October 1971 the first CT scan of a patient with a brain cyst was performed at Atkinson Morley Hospital, London. Hounsfield hypothesised that it is possible to determine what is inside an opaque box by taking X-ray measurements at various angles around it to build an image of the object in slices. Prototype scanners were built and tested on human and animal brains in the run up to first use in a human patient.

CT scanning revolutionised diagnostics in medicine and neurosurgery. The availability of high resolution images superior to simple skull X-rays enabled lesions in the brain to be localised prior to surgery in a way that had never before been possible. Air ventriculography (injecting air into the ventricles of the brain followed by skull X-rays to look for deformity of the ventricles as a sign of an intracranial lesion) would become a thing of the past replaced by a non-invasive imaging technique. However, very soon after the discovery of CT scanning another technological advance would enable highly detailed imaging of brain tissue in particular: the advent of Magnetic Resonance Imaging (MRI)

MRI Scanning

In 1971 Raymond Damadian showed that nuclear magnetic relaxation times (the time taken for magnetic dipoles of protons to resume their natural orientation after being exposed to a strong magnetic field) differed between tumours and non-pathological tissue *in vivo* (Damadian, 1971). In the meantime, Peter Mansfield developed the echo-planar imaging (EPI) method that shortened the time needed to scan to seconds rather than hours. The first MRI of a human being was performed on the 3rd of July 1977 (Hinshaw et al., 1977) heralding the age of clinical MRI scanning. Images of the human brain were of much higher resolution than CT scanning resulting in the use of MRI as the diagnostic imaging modality of choice for the nervous system.

As computational power and magnet power increased the quality of MRI images improved further and new MRI sequences continue to be invented by MRI physicists. MRI is now able to investigate fine structural detail in the human brain as well as underlying brain connectivity and function using modalities such as Diffuse Tensor Imaging (DTI) and Blood Oxygenation Level Dependent (BOLD) functional imaging. Methods to detect chemicals in tissue have been developed such as MRI spectroscopy allowing the physiological state of the tissue to be investigated.

All of these advances led to an unprecedented ability of the surgeon to know before the operation where the lesion is and roughly what it might be. The surgeon then used physical landmarks on the skull and then mentally matched those to the plain film images in the operating room lightbox to determine where to fashion the craniotomy for the lesion shown on the MRI. The remaining step in terms of technical advancement was the use of MRI images in the operating theatre updated with real-time stereotactic position of an instrument so the surgeon can achieve stereotactic accuracy based on MRI during surgery. This was the beginning of neuronavigation.

Neuronavigation

Image guided neuronavigation systems provide the ability to accurately visualise surgical targets during operative procedures. Such systems began to be introduced soon after CT and MRI scanning became more widespread in the 1980s. Surgical tools were tracked using active or passive optical systems and the position of the tool tip was represented on a computer screen displaying the MRI dataset. This information is useful in the operating room when planning the skin incision, craniotomy size and location and the surgical trajectory to the target lesion. This advance allows navigation in 3D space on the basis of a pre-operative MRI or CT scan. A limitation of this technology is that it does not take into account brain shift during surgery. Indeed, the further an operation progresses the more inaccurate the system becomes as the position of brain tissue deviates from that seen on the preoperative MRI scan. This is most crucial for deep lesions which are poorly visualised using the operative microscope but are seen easily on MRI.

Since the 1980s there have been very few further technological developments for brain tumour surgery. The mainstay of current brain tumour surgery remains the use of the operating microscope with neuronavigation for lesion localisation and awake craniotomy with direct cortical stimulation or continuous neuromonitoring to identify functioning brain tissue.

1.4. Conclusion

Over time the evolution of technology for brain tumour surgery has allowed better preoperative imaging to plan craniotomy location and size. Better illumination and magnification of the operative field aids surgical resection- the surgeon can see what they are doing and avoid critical structures. Neurosurgery has evolved from a high-risk procedure using minimal technology to a safer, high-tech procedure with well managed risks. A death on the table is a rare event. However, current technologies have their limitations and as we approach 30 years since the last major advance in brain tumour surgery there remains a need to push the boundaries of what is possible to achieve better intraoperative detection of tumour tissue.

2. Introduction

2.1. Overview

The previous chapter has outlined my personal reasons for embarking on this work and the historical development of surgery for brain tumours. In this chapter a review of primary brain tumours will show how devastating and complex these tumours are and how only recently the underlying molecular mechanisms have begun to be elucidated. The first step in treating these tumours is often surgery and so the difficulties of surgical resection and the limitations of current practice will be discussed. The powerful case of need for systems that can both aid the surgeon in detecting residual tumour and give diagnostic information at the time of surgery will be made. Mass spectrometry and Raman Spectroscopy will then be introduced as methods of chemical analysis with long traditions in the laboratory. The concept of precision oncological surgery will be introduced, with molecular profiling of tissue as the starting point for delivering the right treatment to the right patient at the right time. A hypothesis that rapid, real time and label free molecular analysis is possible in the operating theatre and this could form the basis of precision medicine as applied to neurosurgery will be formed and a thesis outline will then follow laying out the approach to the problem of instant analysis of tissue during surgery.

2.2. The Molecular Basis of Brain Tumours

Molecular markers were first recommended for diagnosis by the World Health Organisation (WHO) only as recently as 2016. This highlights the relatively underdeveloped state of molecular diagnostics in brain tumours when compared to other cancers. Brain tumours stand out for a number of reasons. Human gliomas represent a poorly understood heterogeneous group of brain tumours with a lack of effective treatments. In spite of its importance, rapid *in vivo* molecular profiling of gliomas to investigate targeted therapies has not been possible until the work done in this thesis. In this section the epidemiology and aetiology of brain tumours will be discussed to gain a better understanding of tumour biology. The concept of clonal evolution, treatment resistance and disease recurrence will be examined as the final step in the natural history of high grade brain cancers.

2.2.1. Epidemiology and aetiology

According to the largest tumour database in the world (CBTRUS: Central Brain Tumor Registry of the United States) brain tumours account for 2% of all cancers and yet account for more years of life lost than any other cancer (Ostrom et al., 2015). The overall survival from malignant brain tumours has been unchanged for the past 50 years. The 5 year survival rate for malignant brain tumours is 34 percent, with anaplastic gliomas and glioblastomas having a 5 year survival of 28 and 5 percent respectively. Malignant primary brain tumours remain the most deadly

cancer and can affect people of any age, gender, social and racial class: these cancers can and do affect anyone, anywhere.

Intrinsic brain tumours have been diagnosed on the basis of the dominant cell type on histopathological analysis. The tumour is named after the cell type which is most closely resembled by the dominant cells on histopathology. However, although it has been assumed that this is the originating cell type direct evidence has never existed that this is the case. Indeed, recent evidence suggests that when oligodendrocyte precursor cells are labelled they can be shown to be the cell of origin for gliomas associated with NF1 mutations which are often pilocytic astrocytomas, optic nerve gliomas and glioblastomas and not oligodendrogliomas (Alderton, 2011). This highlights the fact that precursor cells may undergo a number of morphological changes during tumourigenesis and resemble a completely different cell type in the mature tumour. It therefore seems that tumour precursor cells give rise to brain tumours, as found in animal models.

The Cancer Genome Atlas (TCGA) has facilitated a molecular revolution in our understanding of this disease from an imprecise, morphological description with limited reproducibility to a more robust molecular description as embodied in the 2016 World Health Organisation classification of brain tumours (Louis et al., 2016). Characterisation of gliomas now requires identification of isocitrate dehydrogenase (IDH) mutation. Other key changes such as Alpha Thalassemia/Mental Retardation Syndrome X-linked (ATRX) loss and O6-methylguanine DNA methyltransferase (MGMT) promotor methylation are thought to drive gliomagenesis, confer resistance to treatment and have prognostic and diagnostic value when identified (Verhaak et al., 2010), (McLendon et al., 2008).

The new role of molecular analysis in the WHO classification for brain tumours is illustrated in Figure 1. 10 integrated diagnostic groups have been created.

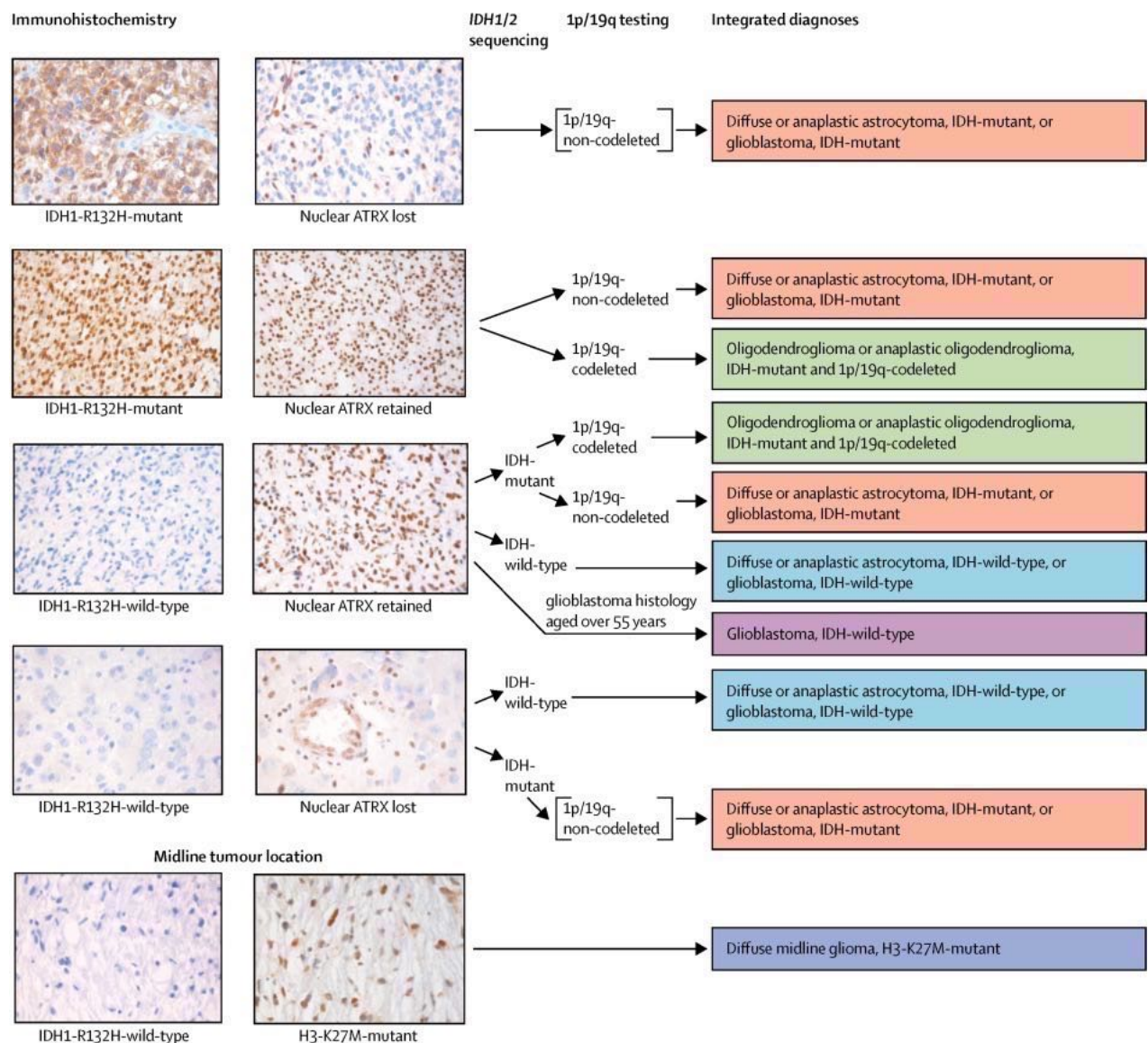


Figure 1. Flow chart of the new diagnoses integrating histological and molecular information in the new WHO classification of CNS tumours.

The integrated diagnostic groups are linked almost directly to treatment modalities, so that the diagnosis almost acts as a prescription for further treatment. The treatment options are shown in Figure 2 and consist of surgery, chemotherapy and/or radiotherapy.

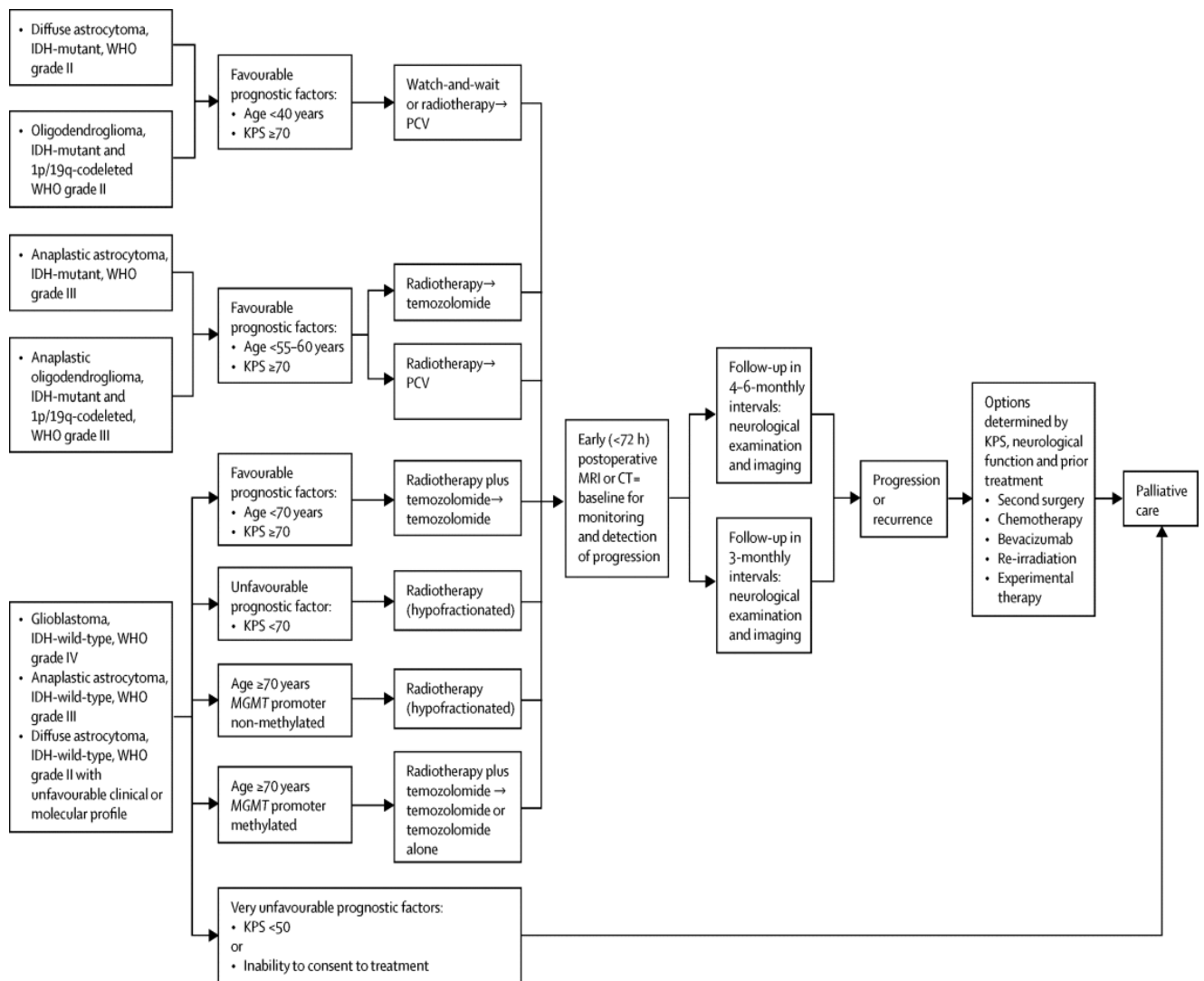


Figure 2. Diagnosis and treatment guideline pathway from the European Association for Neuro-Oncology (EANO)

Glioblastoma (GBM) is a high grade brain tumour and was the first tumour to be analysed in TCGA as it was a model of an aggressive disease lacking effective treatment options. Verhaak (Verhaak et al., 2010) used iterative consensus clustering on 202 GBM samples as part of TCGA. He identified 3 molecular subtypes of GBM:

Proneural (IDH1 and TP53 mutant)
Classical (EGFR mutant)
Mesenchymal (NF1 mutant)

This work was seminal in identifying the molecular basis of GBM being a heterogenous disease with clear subtypes of tumour and outcomes rather than a single disease with a uniformly bad outcome. Further work done by the TCGA Low Grade Glioma network on 293 grade 2 and 3 tumours identified molecular subclasses of low grade glioma (Cancer Genome Atlas Research et al., 2015). 'Clustering of clusters' analysis showed 3 clear groups of low grade gliomas with different survival:

IDH WT LGG with survival of 1.7 years (like GBM)
IDH MT, 1p 19q codeleted LGG with survival of 8 years
IDH MT glioma with survival of 6.3 years

Tumour grade and histology was found to have little effect on DNA methylation pattern with IDH and 1p19q status being the main drivers of clustering. This has encouraged a glioma-wide analysis of molecular features and has in effect redefined glioma as a disease on multiple scales: genetic, molecular, histological and clinical behaviour.

2.2.2. Clones and behaviour

The behaviour of a tumour during treatment and the eventual recurrence of disease in high grade tumours has been a focus of research into improving patient survival. 3 hypotheses have been brought forward to explain the clonal evolution of cancer-branched evolution, big bang and the multiverse hypotheses.

Branched evolution of clones (Sottoriva et al., 2013) is thought to create intratumoural heterogeneity which is the basis of treatment failure. Sottoriva et al. collected spatially distinct tumour samples from 11 patients with GBM showing differing GBM subtypes within the same tumour. Phylogeny constructed from the fragments identified early and late events in the progression of cancer. An alternate hypothesis is the big bang theory (Sottoriva et al., 2015) as exemplified in colorectal cancer. In this hypothesis cells share clonal and sub clonal alterations - therefore, genetic events occur in the tumour field across all cells. The multiverse hypothesis (Lee et al., 2017) purports that greater private genetic events occur at the clinical level, with key early truncal events giving rise to distinct mutational profiles in different areas of the tumour, with minimal overlap and sharing of key mutations. Targeting truncal events with a precision medicine approach is therefore more likely to result in successful control of certain clonal populations.

Integrating this advancing knowledge of clonal evolution into clinical treatment pathways is challenging. A general schema is shown below from the molecular evolution of gliomas working group (Aldape et al., 2018).

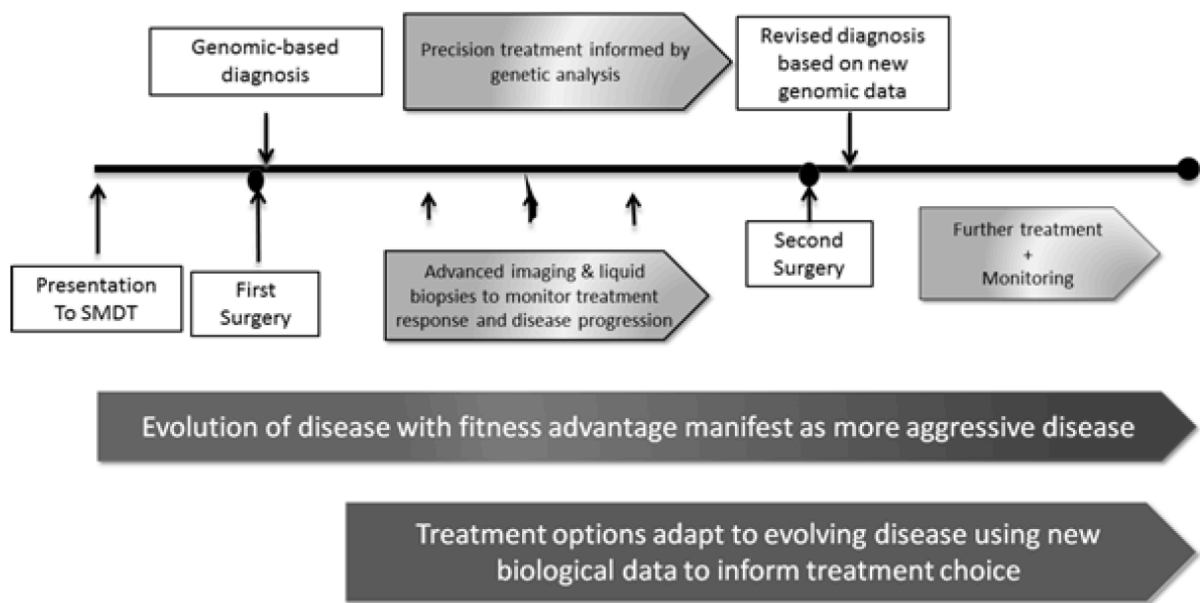


Figure 3. Integrating the molecular evolution of glioma throughout the treatment pathway. (Aldape et al., 2018).

This emphasises the modern approach to glioma. Surgery is required for the initial diagnosis resulting in targeted precision treatment depending on the key driver mutations identified, but tissue sampling is also recommended after the first round of treatment due to clonal selection and a shift in the genetic make-up of the tumour. This may then inform further targeted treatments to control recurrence as the molecular nature of the tumour changes in response to treatment. This is a more precise, personalised approach to glioma care rather than the current simple algorithmic ‘one size fits all’ approach.

2.3. Neuro-Oncological Surgery: Current Practice

The ability to successfully identify and accurately localise the tissue being operated on is one of the determining factors which predict whether the tissue can be removed completely - if the surgeon is unable to identify that which needs to be removed then by definition the surgeon cannot be sure that all of the tissue which could be removed has been removed. According to the analogy in section 1.2 above, one must be able to tell the difference between the grape and jelly if one is to remove the whole grape and leave the jelly relatively untouched. The other aspect to brain tumour surgery is identifying functional brain tissue which cannot be removed even if tumour is infiltrating it. In such circumstances, residual tumour is left behind in the knowledge that damaging brain tissue and creating a permanent neurological deficit precludes the patient from further treatment and in many circumstances results a shorter overall survival with poor quality of life from disability. Characterisation of normal versus tumour tissue is therefore a central task during brain tumour surgery.

2.3.1. Functional-oncological balance

Over time, evidence supporting the maximal Extent Of Resection (EOR) for low and high grade gliomas has been building. The current concept of safe, maximal resection has been championed by a number of surgeons including Mitchel Berger, Hughes Duffau and Walter Stummer.

The concept of maximal EOR has been through a rocky road. At the turn of the 19th century Walter Dandy performed one of the most extreme surgeries possible: removing half of the brain (a hemispherectomy) in his valiant attempts at maximal resection for malignant brain tumours (Dandy, 1928). The tumour still occurred in the contralateral hemisphere. This experience foreshadowed the surgical approach over the next few decades and surgery for GBM was often minimal and accompanied with nihilism. This status quo was challenged in the next century by Lacroix (Lacroix et al., 2001) who established the 'all or nothing' principle. His analysis of over 400 GBM cases suggested that 4 months extra survival benefit was present if more than 98% of tumour volume was removed compared to less than 98% resection. Further studies emerged using volumetric data showing that the relationship between volume of resection and survival was in fact quantitative. Sanai found resection of over 78% volume resulted in survival benefit (Sanai et al., 2011). Probabilistic modelling using data from over 700 GBM patients has shown an almost linear relationship between EOR and survival (Marko et al., 2014) which further suggests that patients do better with more tumour removed.

However important factors that impact the EOR also have an influence on survival. Patient factors (age, performance status) and tumour factors (distance from eloquent area, tumour grade and biological behaviour) can bias results in uncontrolled trials. It is not surprising that in a deep seated high grade tumour in an elderly patient who is already disabled, the neurosurgeon is unlikely to achieve a gross total resection, and due to tumour biology and patient factors survival will be poor. In these respects, EOR may be seen as a surrogate for other factors rather than an independent factor in itself. Only a well designed, controlled trial with enough patient numbers will allow these factors to be teased apart.

There is also debate as to how to measure tumour volume to calculate EOR. For high grade tumours the contrast enhancement on MRI is often used. This may significantly underestimate the true volume of disease as it is well known that histologically, even from the experience of Dandy in the 1920s, that the disease has likely spread well beyond the area of contrast enhancement. For low grade gliomas the tumour can be defined according to the area of T2 FLAIR abnormality. This can underestimate the volume of disease in oligodendrogliomas which exhibit satellite lesions beyond the main tumour bulk. This underestimation of tumour extent is clinically important as tumour recurrence often happens from residual tumour at the margin of previous surgery where disease has been left behind. Indeed, residual tumour volume is an independent prognostic indicator regardless of therapy status, patient age or MGMT status (Ellingson et al, 2018).

A more contentious issue is achieving a high EOR for low grade tumours in an attempt to improve survival by reducing the likelihood for malignant transformation to a higher grade, fatal tumour. Duffau's data on supramaximal resection of low grade

gliomas supports this (Mandonnet and Duffau, 2017), as does the recent Norwegian study comparing biopsy of LGG with resection (Jakola et al., 2012). However conservative options are still often used, ranging from MRI surveillance scans to biopsy only for managing this condition with some only advocating surgery once the lesion is growing and exhibiting high grade behaviour using the rationale that further surgery/chemoradiotherapy can be offered. Other surgeons argue that this is now too late as the patient succumbs to the dismal prognosis of a high grade glioma. The reality is that current practice is somewhere in between these two extremes. An option is needed for surgeons wishing to excise more tumour, but not inflict potential morbidity or undertake the extreme supramaximal surgery proposed by Duffau.

Figure 4 summarises the fact that enough tissue must be removed to improve survival, provide an accurate diagnosis and to improve symptoms of mass effect and yet caution should be exercised when aggressively removing tissue to preserve functioning adjacent brain tissue. The perfect balance would avoid the need for re-operation if non-diagnostic tissue has been removed (sampling error if too limited a sample of tumour tissue is removed, for example). It would also prevent re-operation if a suboptimal volume of tissue was removed at the first sitting when aggressive resection was required. Fundamentally, the surgical task is one of tissue characterisation. The surgeon has to characterise and distinguish tumour tissue from normal brain tissue and make a series of surgical decisions with regards to which parts of the tumour can be removed. The tools available to assist the surgeon in this task are limited but have evolved over the last few decades.

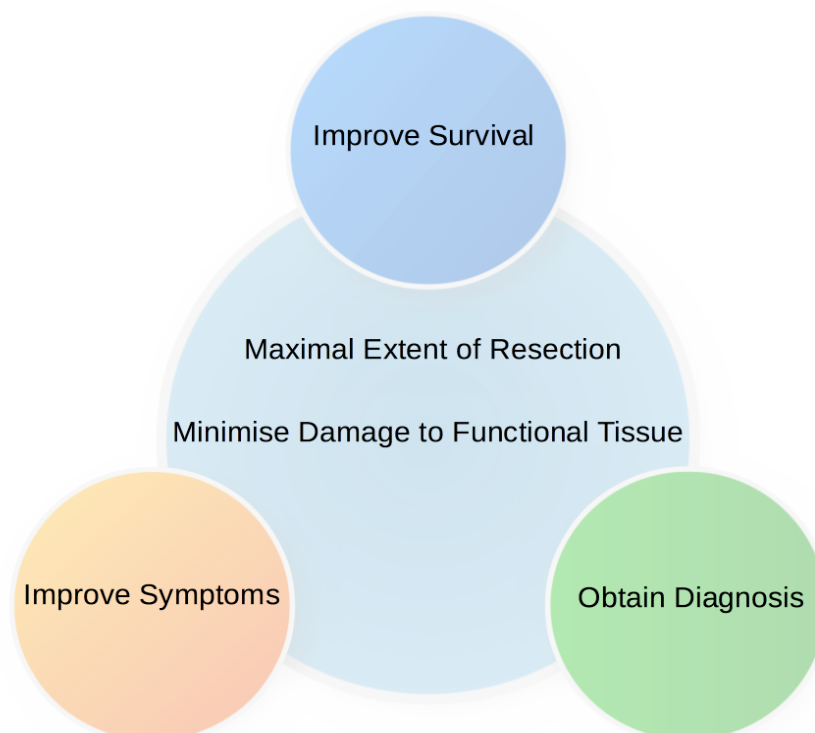


Figure 4. Goals of Neuro-Oncological Surgery.

The 3 goals of neuro-oncological surgery are shown. Each goal consists of balancing resection with preserving neurological function: these are the central concerns during tumour surgery.

2.3.2. Tissue characterisation

Current practice in low grade gliomas has been reviewed in Europe in the largest recent survey by the European Low Grade Glioma (ELGG) Network (Mandonnet et al., 2017). This survey reveals that preoperative imaging and awake craniotomy (where the patient is woken up during surgery and areas of the brain are tested directly by electrical stimulation) are currently used to guide surgery with little emphasis placed on molecular markers despite the central importance in diagnosis, prognosis and as predictors of therapy response. The survey also identifies 2 schools of practice in Europe - those centres performing awake craniotomy for non-language dominant tumours and those not performing awake procedures for such tumours. A lack of consistency in intraoperative functional testing paradigms is also revealed, making it difficult to compare outcomes from awake craniotomy without bias.

The most basic approach to tumour tissue characterisation is that developed since the turn of the 19th century by Cushing, Dandy and more recently by Yasargil: that of mechanical haptic feedback and ‘feel’ of the tumour tissue and the brain/tumour interface (dissection plane) under optimal optical illumination. This can be done macroscopically and microscopically using micro instruments and suction. Over time the feel of tumour tissue as being slightly more rubbery than normal brain tissue develops, and the surgeon can use this as a clue to demarcate the tumour with the use of high magnification through the microscope to optically detect subtle variations in tissue colour and vascularity. Although forming an important part of the ‘art’ of neurosurgery this skill is difficult to quantify, teach and assess between surgeons and more objective methods have been sought not to supplant this skill but to add to it and make it more refined in order to use it only when necessary to demarcate tumour from normal tissue. Medical imaging has been a natural first stop, as suggested by the ELGG survey above.

The imaging methods used to characterise tumour tissue include pre operative imaging using Computed Tomography (CT) or Magnetic Resonance Imaging (MRI) and intra operative imaging using MRI or ultrasound (US). Imaging performed before surgery is a static picture of the tumour at one point in time. CT scanning provides a rapid but low resolution picture of the tumour and contrast enhancement, whereas MRI scanning provides a more detailed picture of the tumour and surrounding anatomy. The preoperative imaging is loaded onto a neuronavigation platform such that any point in the operative field can be matched to the preoperative MRI using a pointer tracked by infrared cameras in 3D space. The problem with preoperative imaging is that the accuracy of the image in terms of representing structures being operated on reduces as the operation progresses and brain shift occurs. The effect can result in loss of accuracy of several centimetres (Barone et al., 2014) which severely compromises the use of preoperative imaging to help define tumour tissue. Intraoperative MRI (ioMRI) is a solution to this as it offers the chance to navigate on updated imaging information during surgery. In practice the use of ioMRI is limited to only those centres which can afford it, and even in those centres it may only be used once at the end of surgery. It is very costly and adds significant time to the operation

as the patient must be transferred to the MRI scanner whilst maintaining sterility (redraping is necessary if sterility is compromised) following which the MRI sequences are then obtained. Intraoperative ultrasound offers a cheaper solution which can be utilised more often during surgery and offers the opportunity to actually guide progressive tumour resection but is plagued by inter-observer variability in the use of the system and interpretation of images.

Due to the drawbacks of conventional imaging modalities other methods have been developed to detect tumour. Optical fluorescence using 5-ALA was first reviewed by Stummer in 2003 (Stummer et al., 2003) and led to the randomised controlled trial in 2006 (Stummer et al., 2006) showing the benefits of fluorescence guided resection in improving survival in GBM. It is not currently uniformly used in the UK and is not currently standard of practice (although may be soon). There are concerns regarding the side effects of 5-ALA administration in terms of photophobia and skin irritation. It is also the practical experience of many surgeons that more than just the tumour fluoresces, with peritumoural reactive gliosis and normal functional tissue fluorescing, hence increased morbidity when pink fluorescence is chased aggressively (Hervey-Jumper and Berger, 2016). Indeed, non-fluorescing tissue can also contain dense tumour as shown in multiple 5-ALA biopsy studies (Lau et al., 2016), (Hauser et al., 2016), (Widhalm et al., 2013), (Hefti et al., 2008), (Coburger et al., 2014)). 5-ALA does not fluoresce low grade gliomas which is the only group of tumours with a large body of evidence and professional opinion advocating maximal volume resection.

Given the benefits and drawbacks of different methods of tumour resection there has been a need for guidelines to standardise the care offered.

2.3.3. Guidelines

In 2006 Improving Outcomes Guidance (IOG) was issued by the National Institute for Clinical Excellence (NICE) CSG10 which set the framework for MDTs and peer review and access to the neuro-oncology MDT. This set of guidance outlined service organisation and a commitment to offer highly organised care from referral to surgery and further treatment. Outcomes from surgery are assessed at the MDT and if the patient is functionally worse as a result of surgery, further treatment may not be offered resulting in worse prognosis than if surgery wasn't offered at all.

In terms of surgical approaches, the guidance mentioned that in LGG, data is lacking and expert opinion on retrospective studies is the best evidence and an association was recognised between EOR and survival. For High Grade Glioma (HGG) the guidance concluded 'one rigorous review concluding insufficient evidence to conclude whether better outcomes arise for HGG following surgical resection vs stereotactic biopsy as only one underpowered RCT was identified'. Although it was recognised that observational studies suggest an association between EOR and prolonged survival, EOR was not recommended as a factor to be measured and postoperative imaging was not recommended.

IOG has been implemented in the UK and has resulted in subspecialisation of neuro-oncological surgery in the vast majority of units and a network of peer reviewed

MDTs. A new guideline is currently in development by the National Guideline Alliance. However crucial questions about EOR and postoperative imaging remained unanswered resulting in significant variation in clinical practice in the UK.

The James Lind Alliance (JLA) placed EOR as one of the top 10 questions in neuro-oncology as part of a priority setting review for the National Institute of Health Research (NIHR) (<http://www.jla.nihr.ac.uk/priority-setting-partnerships/neuro-oncology/top-10-priorities/>)

The London Cancer Brain and Spine Pathway Board: Neuro-Oncology Guidelines are more prescriptive and state that for HGG all patients should be considered for surgery, pathological confirmation required prior to RT/Chemo, complete resection of tumour as safely as possible and biopsy if excision likely to result in major functional impairment if of sufficient performance status. Preop and 48 hour postop (HGG) and 24 postop (medulloblastoma) MRI scans are required to assess the degree of resection and act as a baseline and for use in RT planning. For LGG surgery is seen as important first management for diagnosis. Also extensive resection linked to improving survival - hence maximal safe EOR required.

European ESMO guidelines published in 2014 support maximal safe resection for HGG based on Level 2c evidence (Lacroix et al., 2001) and suggest radical surgery is the backbone of LGG management for diagnostic and therapeutic purposes on level 3b evidence.

EANO guidelines published in 2017 for HGG support maximised EOR while minimising risk of new neurological deficits. Prevention of new deficits positioned as more important than EOR. Standard of care (RANO Neurosurgery 2012) is MRI imaging in the early postoperative phase within 24-72 hours of surgery. Large residual tumour volume is recognised as a negative predictive factor for survival.

AANS/CNS joint guidelines for newly diagnosed glioblastoma (2009) and LGG (2015) in neuro-oncology recommend maximal safe resection. LGG guidelines recommend, based on level 2 evidence, Gross Total Resection (GTR) or Subtotal Resection over biopsy for improved Progression Free Survival (PFS), Level 3 evidence that greater EOR leads to greater Overall Survival (OS). Surgical adjuncts to improve EOR such as ioMRI were recommended, as well as intraoperative brain mapping to improve safety of surgery for eloquent region LGGs. HGG guideline is old but has a level 2 recommendation that maximal safe resection be attempted for newly diagnosed HGG. A level 3 recommendation is the use of biopsy, partial or gross total resection depending on the condition of the patient, size and location of HGG.

Overall the guidelines recommend increased EOR with some variation in older guidelines about early postoperative scanning (there is no UK national guidance for maximal resection, postop scans or EOR measurement). The UK guidelines are currently being revised to address variation in practice.

There is no consensus on what to do with residual disease on immediate postoperative MRI in HGG and LGG. There is also no consensus in any guidelines for which adjunctive technology to use (iMRI, iUS, 5ALA or intraoperative mapping).

2.3.4. Current practice

Surgery is commonly always performed with neuronavigation and microscope. Tissue diagnosis during surgery is commonly performed using frozen section taking 30-40mins which is time consuming, lacks spatial resolution is resource intensive and therefore not repeated during surgery. Molecular diagnosis is not given. The use of adjuncts to surgery and accurate estimation of residual tumour are highly variable due to lack of consensus in guidelines.

In order to capture variation in current practice a number of surveys have been conducted looking at current practice. A survey looking at residual enhancing disease covering 18 Neurosurgery units (Watts, 2016) revealed access to resources for early postoperative MRI and intraoperative technology (iMRI and US) to detect residual disease was variable. Early postoperative MRI to assess volume of resection and residual tumour is not a standard of care yet in the UK.

The RED-GB national service evaluation survey results (MA et al., 2017) shows GTR not achieved in a number of cases. Residual disease is not operated on as there is a lack of perceived clinical benefit. 113 patients with suspected GBM from 15 units undergoing craniotomy for GTR studied. Not all received postoperative scans. GTR was felt to be achieved at surgery in 80% but actually only achieved in 45% of cases.

Why is there is an underestimated amount of significant residual tumour left after surgery? Tissue resection is most difficult towards the end of surgery, when by definition one is most likely to leave residual tumour. This is for a number of reasons:

- Most brain shift at this point, navigation is inaccurate
- Large resection cavity with loss of tissue planes, visually and physically indistinct tumour tissue at the base of the cavity representing tumour origin
- Proximity of functioning brain tissue/at the resection margin limiting further surgery.

At this point a system for rapid identification of tissue type would be of immense benefit as quality control for the operating surgeon wishing to resect as much volume of tissue as safely as possible. Molecular characterization of tissue types offers a highly accurate and reproducible model which such an application demands. The other important demand of the surgeon is that the system must be fast to offer an improvement on the time taken for a frozen section. It must also offer better spatial resolution than frozen section for it to be of use in guiding the resection of residual tumour. The characteristics of the ideal system for tissue identification would therefore be that the system is:

- Rapid (real time)
- Accurate and Reliable (sensitive and specific)
- High spatial resolution (tracked with neuronavigation)
- Safe, easy to use, non destructive

2.3.5. Summary

The choice of adjunct used on top of routine neuronavigation and the operative microscope depends on sensitivity, resolution, cost and additional surgical time. The different types of adjunct can be classified into tumour detection and function detection. Ideally a multimodal balanced approach combining both types of adjunct is required, tailored to the individual patient.

Intraoperative MRI is good for correcting brain shift after craniotomy has been performed but the use of iMRI is scarce as the cost of such setup is high and requires specialised operating rooms including shielding and use of non-ferrous instrumentation. In addition, restriction in the patient positioning makes it difficult to operate on occipital and posterior fossa lesions.

Intraoperative ultrasound is attractive due to low cost, is as accurate as ioMRI in some cases but can be operator dependent and only 1 system is integrated with neuronavigation (Sonowand).

5-ALA fluorescence requires timing of medication, FDA approval, is expensive, has lack of usability for LGG and variability in fluorescence in HGG and hence possible to have ioMRI enhancing non-fluorescing tumour which can be missed.

Awake craniotomy has the drawbacks of time cost, patient anxiety, lack of accuracy with patient tiredness and a lack of accepted paradigm for testing function of specific brain areas.

2.4. Precision Medicine and Machine Learning in Neurosurgery

“It’s far more important to know what person has the disease rather than what disease the person has.” – Hippocrates

Glioma demonstrates the failure of traditional paradigm of biology and clinical trials to develop effective treatments for complex diseases. Although this paradigm worked well for some diseases, for example the use of aspirin for stroke, this type of medicine does not have much to do with each patient specifically and the diagnosis is a general one shared by many other patients and the treatments offered to the whole group are pre-constructed and impersonal.

The traditional paradigm has been a quest for a single target molecule for which a single agent is designed. Similar to traditional neuroscience which started with studying a single neuron rather than massive parallel networks. Traditional clinical trials have also been run along the lines of the same intervention to all members of the test group with placebo/current best care/alternative intervention to control group. For rare cancers such as glioma there has been difficulty in recruiting enough numbers and in identifying subpopulations most likely to benefit as we lack the tools

to undertake this analysis prospectively. The result has been a failure to demonstrate effectiveness of any of new treatments, continued low survival and the continued one size fits all approach (status quo) of chemoradiotherapy for all high grade gliomas.

Cancer is however a mathematically complex non-linear multivariate system. Systems biology has developed the tools necessary to tackle such complex problems using multivariate statistics and machine learning. This has enabled holistic molecular profiling of diseases and disease states using complex datasets which previously would have been difficult to compute. This approach can directly lead to improving cancer care through 2 steps:

- Step 1: New molecular strategies consisting of chemometrics at the point of care for profiling cancer.
- Step 2: New trial designs adapting to molecular profiles of patients and selecting a range of available agents (including repurposing of drugs). This allows significant findings to be recognised in a smaller number of patients, speeding up research. This may also allow individual treatments to be selected and developed and also combinations of treatments to be identified and investigated tailored to the specific patient (personalised medicine).

2.4.1. Definition

Personalised medicine as a concept developed initially following the human genome project and was closely associated with pharmacogenetics. The first relevant PubMed appearance was in 2000 where a BMJ report (Gottlieb, 2000) recognised the importance of a study reporting that specific SNP's determined the responsiveness of asthmatics to inhaled Salbutamol (Drysdale et al., 2000).

“Precision medicine” – sometimes referred to as “individualised medicine,” “personalized medicine,” or “targeted therapy” – is, in basic terms, matching the right patients to the right medicines at the right time. This means gaining a deeper understanding of the specific molecular characteristics that are driving a patient's tumour growth and finding the right treatments to target those specific molecular abnormalities that are responsible for the disease. As an informatics problem, it is complex. Non-linear relationships exist between multiple factors in dynamic systems with enough redundancy so that single pathways and treatments blocking those pathways will never be effective.

The term ‘Precision Medicine’ was first used in a report from the National Research Council in the USA in 2011 titled ‘Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease’ (Disease, 2011). The focus was very much as a first step on creating new definitions of disease more related to functional molecular and clinical data rather than simple histological structure, as is often the case in cancer. In this way Precision Medicine is more aptly described as the marriage of molecular research and clinical data at the point of care. It is in essence a data integration task with the aim of identifying new diagnostic groups correlating with treatment response or clinical outcome. The result is an accelerated analysis at the point of care combining the diagnostic action with a

therapeutic action allowing a combined intervention to take place (hence the coined term theranostics: therapeutic and diagnostic intervention). In this way the disease taxonomy reflects biological behaviour of the condition being assessed: it will reflect the intrinsic biology of the disease and goes beyond just a simple physical description. This is similar but different to the concept of personalised medicine which relates to the selection of the best treatment for an individual: by analysing diseases into new groups precision medicine offers more than treatment selection but also identifies factors involved in disease prevention, progression as well as response to treatment. Examples currently used in clinical practice include:

Blood grouping and transfusion
BRAF and melanoma
HER2 expression in breast cancer
Oncotype DX testing in breast cancer

To determine the chance of recurrence of breast cancer, the Oncotype Dx® test analyses the activity of 21 genes to yield a highly precise picture of the individual cancer the patient is suffering from, including the likelihood that the breast cancer will return and the likelihood of benefiting from chemotherapy at the early stage of the cancer. Based on the test results, the Oncotype DX assigns a Recurrence Score that reflects the likelihood of the cancer recurring. For example, for early stage invasive breast cancer, scores lower than 18 indicate a low risk of recurrence (suggesting that the benefit of chemotherapy is likely to be small and will not outweigh the risks of side effects); scores of 18-30 reflect an intermediate likelihood of recurrence (risk/benefit ratio unclear); and scores greater than 30 indicating a high risk of recurrence (benefits of chemotherapy likely outweigh the risks).

The exact platform precision medicine takes is variable. Genetics platforms have been described (considered as the ultimate molecular analysis by making use of the sequencing of the human genome), but any form of molecular analysis may be used which provides the dimensions necessary to measure the diagnostic features of interest. The most powerful analysis is arguably a whole system analysis of nucleic acid, proteins and lipids which is possible using Raman spectroscopy and mass spectrometry. It is only recently that the right combination of technologies and methodologies of data analysis have emerged that can enable a precision medicine approach to brain cancer and brain cancer surgery.

2.4.2. Prerequisites

Several advances have underpinned the emergence of precision medicine as a problem solving approach. Advances in data science have been crucial. In particular the widespread development and use of machine learning in statistical packages such as 'R' and mathematical packages such as Matlab has made this approach available to more researchers. The computational power required to drive these machine learning methods has only recently become economical and available - machines with enough central processing unit power and random access memory to carry out billions of operations per second are now widely available. Similarly,

hardware advances in rapid analysis of large amounts of molecular data have enabled the real-time profiling of tumours. This would have been science fiction only years ago but is now a possibility. As a result, there is more patient level molecular data potentially available than ever before.

The emerging characteristics of precision medicine can be summarised as:

- Data driven:
 - Molecular data platform
 - Clinical data platform
- Integration of these molecular and clinical datasets in real-time at the point of care.
- Individualised and personal: utilising individual clinical data with individual-individual variation.

2.4.3. Promise for gliomas

The development of precision medicine in Neurosurgery has been connected to the identification of prognostic, predictive and diagnostic molecular biomarkers. These range from proteins and lipids to genes. Although there are putative biomarkers in multiple areas of neurosurgery, their low sensitivity and specificity preclude regular clinical use in all areas except neuro-oncology. Current biomarkers used in neuro-oncology are relatively new and by no means complete in brain cancer. There is a need for more accurate predictors of:

- Disease growth
- Response to treatment
- Disease outcomes
- Tumour type- new classifications of disease

The main uses of precision medicine approaches to brain tumours have been in tumour detection (Jermyn et al., 2015b) (Kamen et al., 2016) (Santagata et al., 2014) (Ji et al., 2013) (Ritschel et al., 2015) (Kalkanis et al., 2014) and outcome prediction (Akbari et al., 2016), (Azimi et al., 2015), (Emblem et al., 2015), (Yang et al., 2015). Nearly all methods have been based on imaging or *ex vivo* data and have used a variety of methods to collect profiling data, ranging from genomics and imaging analysis to chemometric methods such as mass spectrometry and Raman spectroscopy. The accuracy of these machine learning models has reached 80-90% sensitivity and specificity which is high enough to be useful clinically. Precision medicine also has a role to play for newly emerging treatments such as immunotherapy in glioma. Molecular profiling can help with patient selection for treatment which can be immediately given or given faster for better results, for example DCVax, the dendritic cell vaccine for GBM (Liau et al., 2018). This is a personalised vaccine derived using the patient's own dendritic cells which are trained to recognise the patient's own glioma. First analysis of the trial results in 232 patients shows an increase in overall median survival of 23 months, a substantial increase from the current median survival of 14.5 months.

Precision medicine also has a lot to contribute in terms of the efficient organisation and delivery of services. The Get It Right First Time (GIRFT) 2018 review of neurosurgery services in the UK (Phillips, 2018) was aimed at reducing unwanted variation in service provision. Surgeon led visits analysing data from each neurosurgical centre highlighted the need to:

1. Reduce the proportion of malignant brain tumour cancer patients admitted through emergency pathways through remodelling elective pathways
2. Ensure that diagnostic and therapeutic interventions are undertaken in the right setting including one-stop outpatient facilities

These recommendations are designed to help improve the quality of care and outcomes for patients as well as produce cost efficiency savings.

Increased admission through emergency pathways can occur when elective pathways are already full. This can occur most readily with a one size fits all approach leading to craniotomies for most tumours rather than a select few which will benefit most. Currently we lack the information to be able to select these patients. Other recommendations were that one stop outpatient management of brain tumours should occur in the right setting i.e. as soon as possible after surgery when full diagnostic information is available and further chemoradiotherapy can be planned. The fast provision of diagnostic information as early as possible in the patient pathway is a key enabler of the proposed more streamlined diagnostics and therapeutics and precision medicine has a key contribution to make here. Precision medicine can offer smarter use of existing limited resources enabling more patients to be treated sooner.

2.4.4. New models

Modelling of the human brain is highly complex. Even today with the advent of functional MRI, direct cortical stimulation is still considered more accurate in the localisation of cerebral function in patients with tumours. This is due to false negative findings associated with peritumoural hypervascularity, and false positive findings associated with perilesional oedema (Elder and Rao, 2017). The truth remains that despite our efforts, the best way of modelling the human brain for surgery in eloquent areas is to create hypotheses based on cortical mapping in an *in vivo* model involving an awake craniotomy. In a similar way, *in vivo* mapping of the tumour is essential to understand tumour biology and behaviour in more detail in order to develop more effective treatments. Human glioma is a complex tumour which involves a complex host-disease relationship between tumour tissue and the surrounding brain parenchyma due to the following interactions:

- The blood brain barrier
- Tumour Micro-environment
- Immune system interactions at the brain-tumour interface
- Intratumoural heterogeneity
- Internal genetic drivers determining tumour behaviour

Tumour behaviour can be summarised as having a quiescent phase, slow growth,

rapid growth and necrosis. No one model has been created with accurately replicates this complex tumour system. *in vitro* models involving immortalised cell lines creating tumour spheres lack interactions and checks and balances provided by the in situ microenvironment, and treatments developed using *in vivo* animal models invariably fail when applied to humans. Glioma is a complex learning system able to circumvent anticancer treatment and there has been a failure of the deterministic approach in glioma to offer effective diagnosis/targets/therapies. A holistic, probabilistic approach which looks at multiple scales and systems together may offer a more intuitive and powerful solution. Only recently have the factors preventing *in vivo* holistic glioma modelling and analysis been overcome - namely the problems of in situ data generation and processing.

2.4.5. Summary

Glioma lacks appropriate definition in terms of molecular and genetic determinants of disease. Where routine histology has failed a new modality is required which has high enough dimensionality to profile the biological tissue to the extent that tumour classification is possible. Neurosurgery, despite having navigation and microscope technology, currently lacks molecular data to guide surgery with a reliance on surrogate markers which are often inaccurate and not real time. A novel modality is required to produce this complex dataset. The problem posed by the analysis of complex datasets produced by the new generation of chemometric methods is best solved using newly developed machine learning.

This PhD investigates 2 first-in-man molecular analysis platforms and attempts to integrate the associated molecular and clinical data sets in real time at the point of care in the operating theatre using advanced machine learning models. This represents a novel approach to the complex problem of treating glioma. In the following chapters mass spectrometry and Raman spectroscopy will be introduced as chemometric methods which can form the basis of *in vivo* real time molecular analysis for precision medicine.

2.5. iKnife: Clinical Mass Spectrometry and Precision Medicine

The sections of the thesis that follow have been published in the Handbook of Metabolic Phenotyping: Chapter 7, titled 'The iKnife: development and clinical applications of Rapid Evaporative Ionization Mass Spectrometry'.

Abstract

The application of rapid evaporative ionization mass spectrometry (REIMS) to the development of a new surgical tool, the iKnife, is described. The iKnife, an electrocauterization device linked to a mass spectrometer, provides real-time read-out of the molecular composition of tissue under investigation during surgery. Applications of this novel technology in brain and breast cancer surgery are reviewed. Its extension to endoscopic examinations is described. The use of the technique for characterising microbiological species is also shown. Finally, some thought on the future use of the iKnife are presented.

2.5.1. Introduction

Precision medicine can be defined as "Coupling established clinical–pathological indexes with state-of-the-art molecular profiling to create diagnostic, prognostic, and therapeutic strategies precisely tailored to each patient's requirements" (Mirnezami et al., 2012).

The aim is to provide the right treatment to the right patient at the right time. In the past standard treatments were given to patients having the same condition with poor results in some cases because only a subset of the patients actually responded to the treatment. We now have access to a vast amount of disease-specific molecular information which can vastly improve patient care by identifying previously unknown patterns of disease. Indeed, the World Health Organization has recently published new guidelines on the Classification of Tumours of the Central Nervous System (Louis et al., 2016) which incorporates, for the first time, molecular information for the diagnosis of human brain cancer. However, gene specific information is not enough. A holistic picture, using systems biology, must be captured of the metabolic derangements in cancer to fully profile these diseases and this requires a different form of analysis of tissue to generate the molecular information required to identifying new groups of patients which can respond to new and existing treatments. In short: genomic sequencing is not enough to improve poor cancer outcomes.

The iKnife represents a system which can generate a rich set of molecular information at the point of care- be it in the operating room or the pathology lab- which can be associated with particular disease states. The ability to do this and answer some fundamental questions in clinical care for a range of cancers was demonstrated by Balog et al. (Balog et al., 2013). The most immediate question for the surgeon removing cancer is whether tumour and non-tumour tissue can be easily identified, thus aiding complete resection with minimal damage to normal surrounding structures. One of the most important questions for the pathologist is how to report on the tissue sample obtained by the surgeon to accurately convey the

type of tumour with respect to clinical outcome. For the treating oncologist facing the patient in clinic with the pathology report the main questions are which current treatment should be used for this particular patient, and if recruiting for clinical trials how to stratify the patient for emerging treatment agents. Although these questions have yet to be answered, the development of the iKnife is an important step in the direction of having a platform for precision medicine.

In the following sections of this chapter we aim to explain how iKnife technology was developed, in which clinical areas of research it is being used at present and which future clinical applications may be possible.

2.5.2. Technology

This section outlines the development of surgical mass spectrometry from basic principles to surgical model development.

2.5.2.1. Basic principles of mass spectrometry

Mass spectrometry involves getting the sample of interest into the vacuum environment of the spectrometer in the form of ions suitable for mass analysis. Samples may be treated and pre separated in a number of ways and ionised under ambient or non-ambient conditions depending on the state of matter or phase in which the samples are in. The ions are then collected and fed into the mass spectrometer which then detects the mass and charge of the ions allowing a mass spectrum to be created consisting of mass/charge ratio on the x-axis against signal intensity on the y-axis.

Mass spectrometry has a 50 year history of being used to analyse biological tissue (Snedden and Parker, 1971). Analysis was initially time consuming and difficult owing to the multiple steps involved, the potential for cumulative errors and the need for a substantial amount of mass for adequate signal to noise ratios. Tissue samples were subjected to sample preparation including homogenisation, extraction, chromatographic separation and then mass spectrometry analysis as part of this multi-step procedure.

2.5.2.2. Non-ambient and ambient mass spectrometry

Vacuum conditions had been used previously as a simple choice for creating and maintaining ions for mass spectrometric analysis for liquid chromatography-mass spectrometry (LC-MS), secondary ion mass spectrometry (SIMS) and matrix-assisted laser desorption ionization (MALDI). These techniques require extensive sample preparation and ionization in a vacuum and are still not suitable for rapid clinical use. However, an atmospheric pressure version of MALDI was an important proof of concept that ionisation suitable for mass spectrometry can be performed in air (Laiko et al., 2000).

Ambient mass spectrometry methods were later developed which allow mass spectra to be recorded on samples in their native environment, without sample preparation or pre-separation. These methods include desorption electrospray ionization (DESI), direct analysis in real time (DART) and atmospheric pressure solid analysis probe (ASAP) among others. In the DESI method a fine spray of charged droplets picks up small and large biomolecules, ionises them and delivers them through an atmospheric inlet as desolvated ions into the mass spectrometer. The advantages of ambient methods are that they are fast (taking only a few seconds), reduce the need for sample preparation and integrate well with other methods for well defined chemical species (LC/MS).

2.5.2.3. Development of REIMS (the iKnife)

The Takats group discovered in 2009 that routine surgical instruments which used thermal ablation in the form of electrosurgery or infrared laser ablation produced large numbers of ions from the tissue in gaseous form which are suitable for mass spectrometer analysis (Schäfer et al., 2009). The ionisation of the sample takes place at the surgical site at the time electrocautery is used to cut tissue or stop bleeding through thermocoagulation. High-frequency current is applied to the tissue with the electric energy dissipated in the tissue causing ionisation during tissue evaporation. Plastic tubing connects the electrocautery handpiece to the atmospheric interface of the mass spectrometry machine. In this way gaseous ions are transmitted from the operation site to the mass spectrometry machine for analysis within seconds. The results in terms of tissue specificity were similar to DESI but without any sample preparation (Balog et al., 2010). The atmospheric interface has also undergone modification to help improve signal to noise ratios by removing impurities by passing surgical smoke over a heated coil prior to entry into the mass spectrometry machine. The ionised large and small molecular fragments then come across a collision surface in order to break them down into numerous smaller fragments more appropriate for analysis. These then enter a Quadrupole Time Of Flight (QTOF) mass spectrometry analyser where the time taken for the charged molecular fragments across a specific mass range to travel a fixed distance in a vacuum is measured by a detector. The time taken to travel through this vacuum is a function of the mass to charge ratio for each ionised fragment measured. The resultant data is represented as a mass/charge ratio against signal intensity.

The main advantages of this ambient mass spectrometry method are that it is rapid (analysis takes only a few seconds), requires no sample preparation at all and integrates well with instruments and tools already in use by surgeons with minimal need for modification of surgical methods or instruments. All that is required is for surgical smoke - normally evacuated from the operating room - to be channelled into the atmospheric interface of the mass spectrometry machine for analysis. This has enabled a vast database to be constructed of mass spectra from diseased and normal tissue both *in vivo* and *ex vivo*.

2.5.2.4. Model generation: the concept of profiling

Aside from the hardware advances in mass spectrometry which have enabled the iKnife to be used for the collection of mass spectrometry data during surgery there have been important advances in machine learning which underpin the ability to use the data obtained to answer important clinical questions. For molecular profiling using mass spectrometry a method capable of the parallel determination of thousands of system parameters is required. This needs to be highly reproducible in order to create an authenticated database with the identification of elements by an accepted alternative method. Unknowns are then identified through similarity scores and multivariate statistical models by similarity to known groups in the database.

The process involved in using the iKnife for tissue characterisation in clinical applications requires 3 stages, shown in Figure 5:

1. Database building using *in vivo* and *ex vivo* data
2. Model creation using machine learning
3. Real-time classification (testing the model)

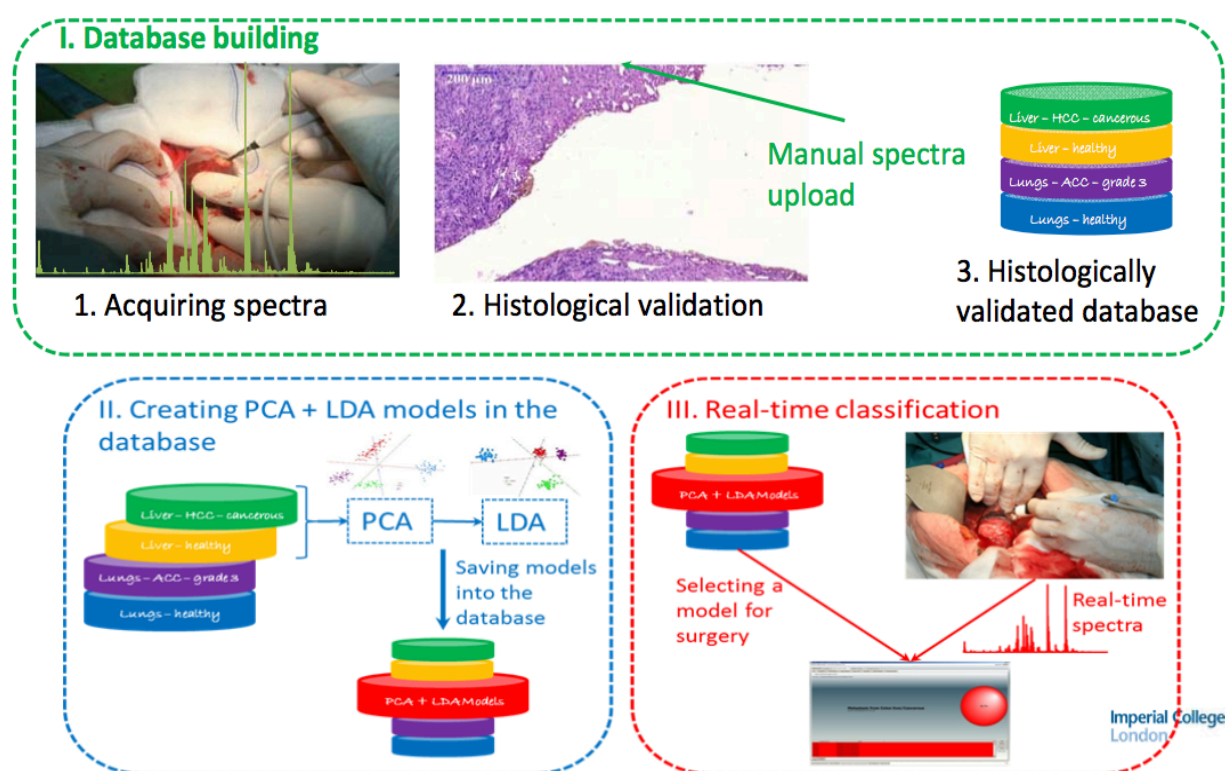


Figure 5. Data flow in molecular profiling of human tumour tissue.

Created by Julia Balog.

2.5.3. Clinical Applications

2.5.3.1. Oncological surgery: introduction

Surgeons play a central role in modern cancer care. Often, the patient pathway starts with the involvement of the surgeon as a vital step in diagnosing cancer by

safely obtaining a physical tissue sample for definitive histopathological analysis. Furthermore, the skills of the surgeon may be called upon to safely remove part or all of the cancer as part of a therapeutic procedure in order to improve symptoms and survival outcomes for patients.

The diagnostic information obtained by the surgeon must be accurate (the sample must be representative of the tumour) and must be obtained without causing functional damage to normal tissue - without direct physical damage or bleeding and haematoma formation. In this context it is essential that the surgeon obtains a sample which is suitable for diagnosis and the surgeon may wait for a frozen section of tissue to confirm this before finishing the operation. Frozen section is a resource intensive and time consuming process with wait times of 30 minutes or more whilst a pathologist reports on the specimen with results not being as accurate as formal fixed histology.

Surgical removal of tumour can be indicated in order to reduce disease burden and improve progression free or overall survival. In such cases it is of paramount importance that whilst every effort is made to remove as much tumour as possible damage to normal tissue is kept to a minimum. In breast cancer surgery this approach can be termed breast conserving surgery (with other treatment modalities applied to treat any residual tumour) although more radical surgery which removes the tumour and a wide margin of surrounding non-tumour tissue can also be performed. Removing a margin of non-tumour tissue around a brain tumour is not possible as surrounding brain function must be preserved. This balance between obtaining a maximal resection of tumour volume but not causing injury by damaging surrounding functioning brain tissue is termed the onco-functional balance (Mandonnet and Duffau, 2017).

The iKnife has a central role to play in helping surgeons attain the 2 main goals of oncological surgery as described above. It is a potential tool for tumour characterisation and diagnosis, aiding or even replacing intraoperative tissue analysis techniques, and is also a potential tool for both intraoperative margin assessment and for guiding the safe, maximal resection of tumours during surgery by its ability to identify tumour and non-tumour tissue in real-time.

The iKnife has been used as part of clinical trials in a number of body organs. Here, we will discuss the use during surgery for breast cancer, brain cancer and colorectal cancer.

2.5.3.2. Breast cancer

Breast cancer is the commonest cancer in women, with the majority of operations in the UK and USA undergoing Breast Conserving Surgery (BCS). Over 20% of patients require re-operation for inadequate margins due to the high risk of relapse. This has important physical, psychological and financial implications for patients and healthcare systems. Although frozen section as a form of margin assessment during surgery has been demonstrated as accurate, its widespread adoption has been limited by the availability of trained personnel, pathology resources, slow turnaround

times and the problems of false positive interpretation. A range of technologies have been developed to provide a tool which offers accurate results during surgery to optimise surgical resection and prevent re-operation (Table 1) with high sensitivity and specificity for detecting the tumour margin.

TABLE 4. Simple Averages of Percentage Sensitivity and Specificity Data of IMA Techniques Compared With Pooled Meta-analysis

IMA Technique	No. of Studies: MA*/All†	Sensitivity			Specificity		
		Pooled MA	Average – MA Studies	Average – All Studies	Pooled MA	Average – MA Studies	Average – All Studies
Total Studies ❖	35❖/55❖	35❖	35❖	55❖	35❖	35❖	55❖
Frozen section	9/12	86%	83.5%	81.0%	96%	93.5%	94.4%
Cytology	11/14	91%	80.9%	82.9%	95%	91.9%	93.5%
Intraoperative ultrasound	4/6	59%	57%	55.7%	81%	79.6%	84.4%
Specimen radiography	9/13	53%	51.6%	54.8%	84%	76.6%	73.9%
Optical spectroscopy	3/6	85%	88.2%	81.7%	87%	81.5%	81.9%
Margin probe	0/4			70.4%			58.5%
MicroCT	0/1			83.3%			94.7%

❖=One study contributed data to 2 different IMA techniques. All indicates all studies included in systematic review. MA indicates meta-analysis.

*Same studies as meta-analysis.

†All studies from systematic review.

Table 1. Simple averages of sensitivity and specificity for IMA techniques in breast cancer surgery. (St John et al., 2017a)

The iKnife offers a novel method for the rapid, accurate characterisation of neoplastic tissue which integrated well with commonly used electrosurgical instruments. The system has been used to analyse electrosurgical vapours during breast surgery to build a histologically validated database of spectra of tumour and normal tissue (Figure 6). When tested against fresh unseen specimens, the model generated was able to successfully achieve 90.9% sensitivity and 98.8% specificity.

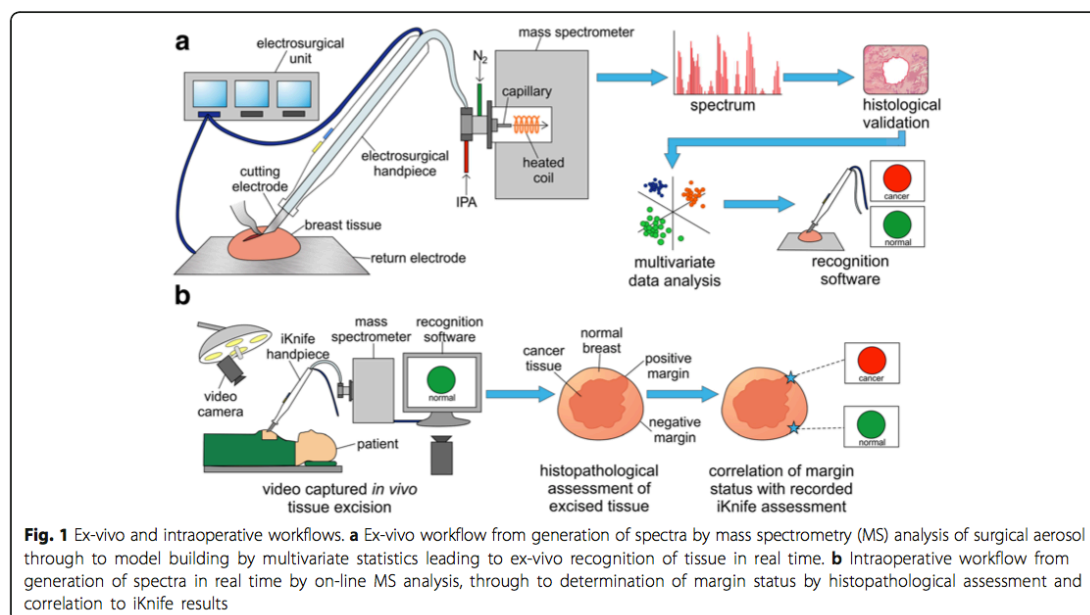


Figure 6. ex vivo and intraoperative workflow: breast cancer iKnife (St John et al., 2017b)

The model when used *in vivo* during surgery was able to interpret 99.3% of spectra

produced during surgery in real-time within 1.80 seconds. MS/MS analysis was used to characterise key lipid species present in cancerous and normal tissue. All of the lipids identified in the Table 2 are glycerophospholipids with phosphatidylethanolamines being the most commonly identified species. The accuracy compares well with existing technologies for margin detection, leading the way for a CRUK-backed multi-site trial of the iKnife as a novel margin detection tool.

Table 2 MS/MS-based phospholipid identification increased in breast cancer (600–850 *m/z*) (*Continued*)

770.57	PC (16:0/20:3)	[M-CH ₃ -H]-
	PE (20:2/18:1)	[M-H]-
	PE (18:0/20:3)	[M-H]-
	PE (18:0/20:2)	[M-H]-
772.58	PC (18:1/18:1)	[M-CH ₃ -H]-
	PC (18:0/18:2)	[M-CH ₃ -H]-
	PE (18:0/20:1)	[M-H]-
	PC (18:0/18:1)	[M-CH ₃ -H]-
792.55	PS (O-18:0/18:2)	[M-H]-
	PE (P-18:1/22:6)	[M-H]-
	PC (16:0/18:2)	[M + Cl]-
	PC (18:2/20:3)	[M-CH ₃ -H]-
794.57	PC (18:0/20:5)	[M-CH ₃ -H]-
	PC (18:1/20:4)	[M-CH ₃ -H]-
	PE (18:1/22:4)	[M-H]-
	PE (18:0/22:5)	[M-H]-
	PC (16:0/22:4)	[M-CH ₃ -H]-
	PC (18:1/20:3)	[M-CH ₃ -H]-
	PC (18:0/20:4)	[M-CH ₃ -H]-
	PC (16:0/18:1)	[M + Cl]-

Possible lipid identifications from tandem mass spectrometry (MS/MS) data for each significant *m/z* peak in negative mode. Numbers within brackets represent the number of carbons in the fatty acid chain, followed by the number of double bonds (C:N). *PA* phosphatidic acid, *PE* phosphatidylethanolamine, *PC* phosphatidylcholine, *PS* phosphatidylserine

Table 2. MS/MS based phospholipid identification in breast cancer. (St John et al., 2017b)

2.5.3.3. Brain tumours

Brain tumours are the leading cause of death for young people. The majority of tumours arising from brain tissue are highly malignant. Gliomas make up 74.6% of all malignant primary brain tumours (Ostrom et al., 2016). Current treatment options for brain tumours include surgery to determine the diagnosis, improve symptoms or survival by removing as much of the tumour as possible, followed by radiotherapy and/or chemotherapy. There is a growing body of evidence suggesting that increased Extent Of Resection (EOR) of brain tumours leads to improved survival

(Brown et al., 2016) (Sanai and Berger, 2008) (Lacroix et al., 2001) (Berger et al., 1994) (Hervey-Jumper and Berger, 2016). Attempts at increased EOR must not result in damaging nearby functioning brain tissue as the resulting neurological deficit may preclude any further treatments resulting in a worse outcome for the patient than if surgery was not performed at all. Unlike in other areas of cancer surgery, taking a surrounding margin of healthy tissue in the brain is not an option as this carries significant risk to the patient, and so the resection must be tailored for each patient. The ability to carry out molecular characterisation of brain tissue in real-time using the iKnife therefore represents an important step in the evolution of safe neurosurgery for the resection of brain tumours.

The iKnife has been modified to work in parallel with the most commonly used surgical tool in brain surgery- bipolar electrocautery forceps. These forceps allow the controlled application of electric current to a small volume of tissue typically under the magnified vision of an operating microscope as shown in Figure 7 below. The resulting smoke is aspirated under a low suction pressure through an opening in the forceps and into plastic tubing connecting to the atmospheric interface of the iKnife. A heated coil removes impurities prior to the smoke entering a spherical collision surface prior to entry into the MS system.

The operating room setup as shown in Figure 7 shows the bipolar method of electrocoagulation as used during surgery. The use of suction bipolar forceps allows the surgical vapour to be sampled directly at the site of use of the forceps and integrates the iKnife directly with surgical use without the need to modify any equipment.

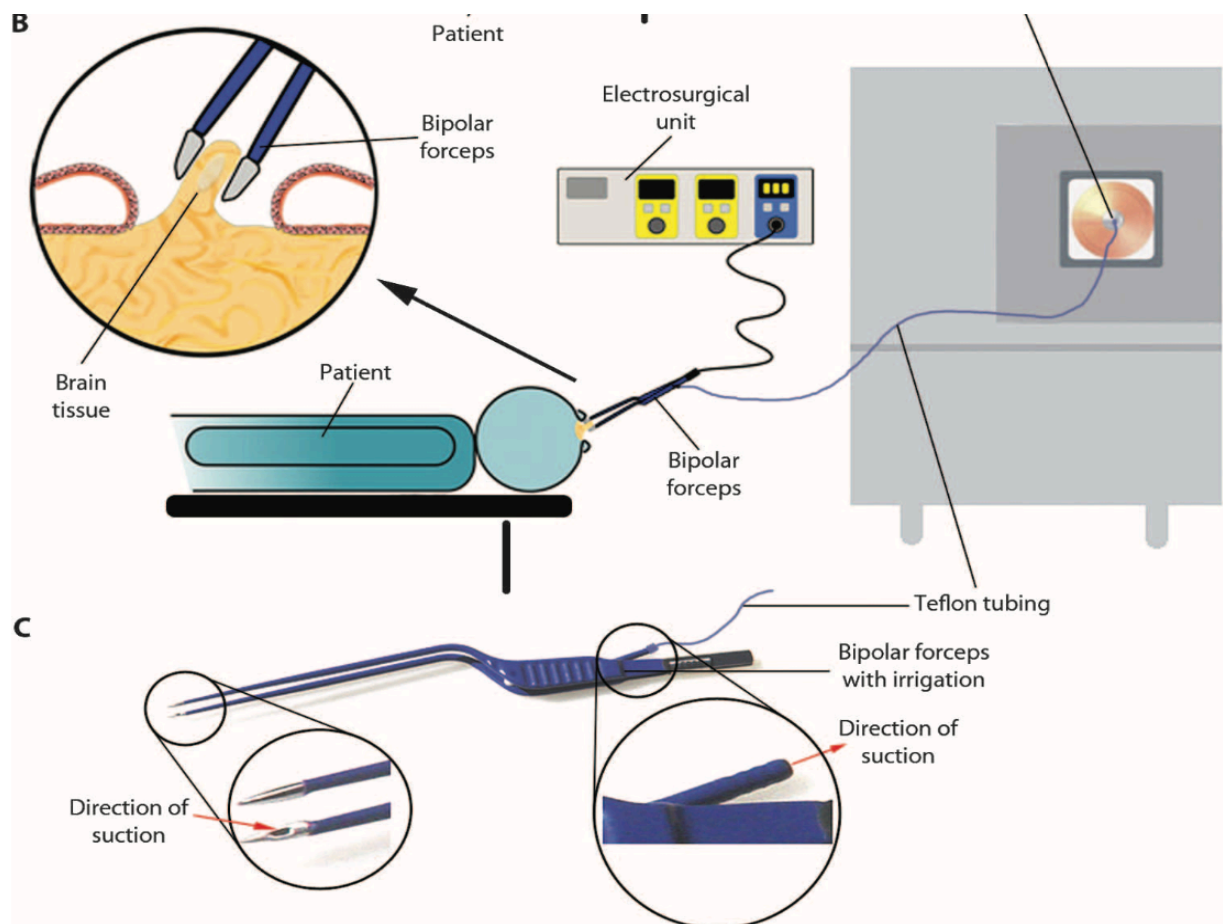


Figure 7. REIMS diagram in the operating room, (Balog et al., 2013)

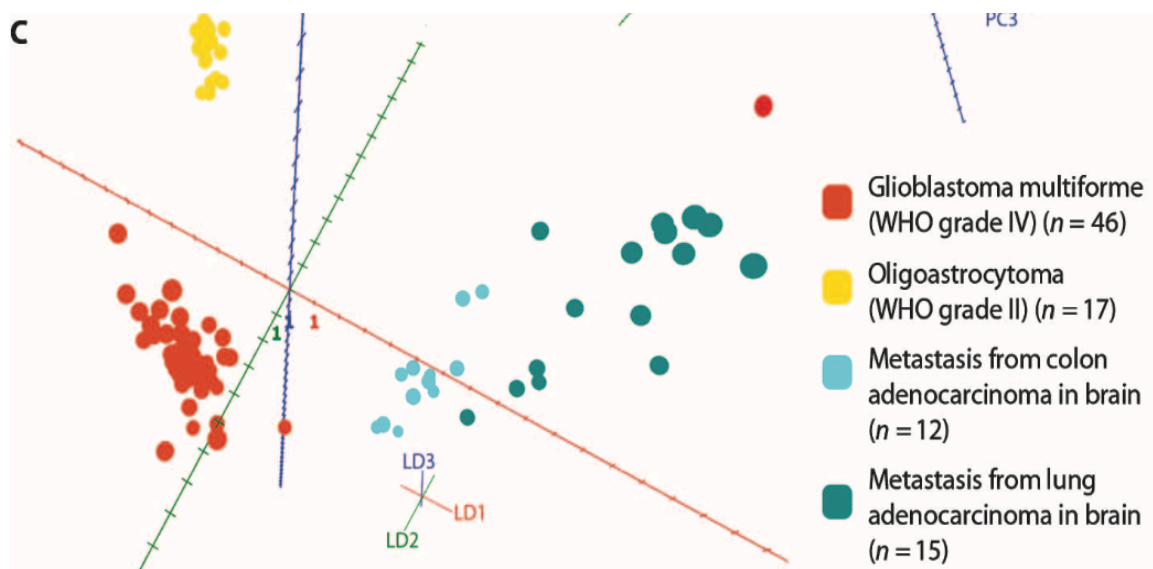


Figure 8. LDA analysis of data from iKnife in a range of tumours. (Balog et al., 2013)

As well as a validated ability to characterise abnormal tissue, the iKnife can also identify clinically important phenotypes at the time of surgery which will help stratify patients into prognostics groups early. The system has been shown to correctly classify 100% of brain tumour spectra with LDA analysis of this dataset showing

excellent separation of disease groups shown in Figure 8. Although currently used as a research tool only, in the future this technology could have important implications for clinical trials of new locally delivered agents, improved survival and quality of life from safer more extensive tumour resection and improved survival from starting conventional treatments earlier on the basis of accurate, immediately available molecular pathology.

2.5.3.4. Gastroenterology

Despite developments in cross-sectional imaging platforms such as CT and MRI, endoscopy is still the gold standard modality for the diagnosis of gastrointestinal disease and is increasingly used as a therapeutic tool for the treatment of lesions such as pre-cancerous polyps in the GI tract. Technological evolution is allowing ever more detailed images of such lesions, but there remains an unmet demand for accurate, objective, user independent assessment of lesions to guide endoscopic and further management (Gupta, 2015) (Rees et al., 2017).

In contrast to the growing stable of optical technologies, the iEndoscope has been developed to provide real time biochemical phenotyping of gastrointestinal pathology during endoscopy based on the mucosal lipidome. As in traditional surgery, electrocautery is routinely used in endoscopy to remove or ablate polyps and other lesions. A snare is passed down the working channel of the endoscope and placed around the base of the polyp. As the snare is closed, safe resection of the lesion is aided by passing an electrical current through the snare wire to heat the tissue and minimise bleeding. As a by-product of this process, smoke is produced within the bowel lumen, containing gaseous ions from the tissue.

The effective transmission of gaseous ions from the site of electrocautery inside the patient back to a mass spectrometer presented a significant challenge. Consequently, a modified snare was developed which incorporated small fenestrations in the snare sheath and a side-port in the handpiece, to permit efficient collection of the gaseous ions (Figure 9).

To date, the iEndoscope has been used as a research tool only. Its feasibility and utility in the upper and lower GI tract was first demonstrated in 2015 (Balog et al., 2015). Subsequent studies on human colorectal tissue have demonstrated that this platform can reliably distinguish between polyps and normal mucosa, as well between cancer and polyps, with overall accuracies of 97.5% and 94.4% respectively (Alexander et al., 2017). Colorectal cancer was characterised by over-expression of long-chain phosphatidylserines and phosphatidylglycerols and polyps demonstrated over-expression of ceramides. Additionally, this study raised the exciting possibility that REIMS might be able to classify cancers according to their histological and molecular phenotype, offering the endoscopist information that hitherto has required lengthy histological analysis.

With the backing of NIHR funding, the iEndoscope is now moving forward into larger scale clinical trials in humans.

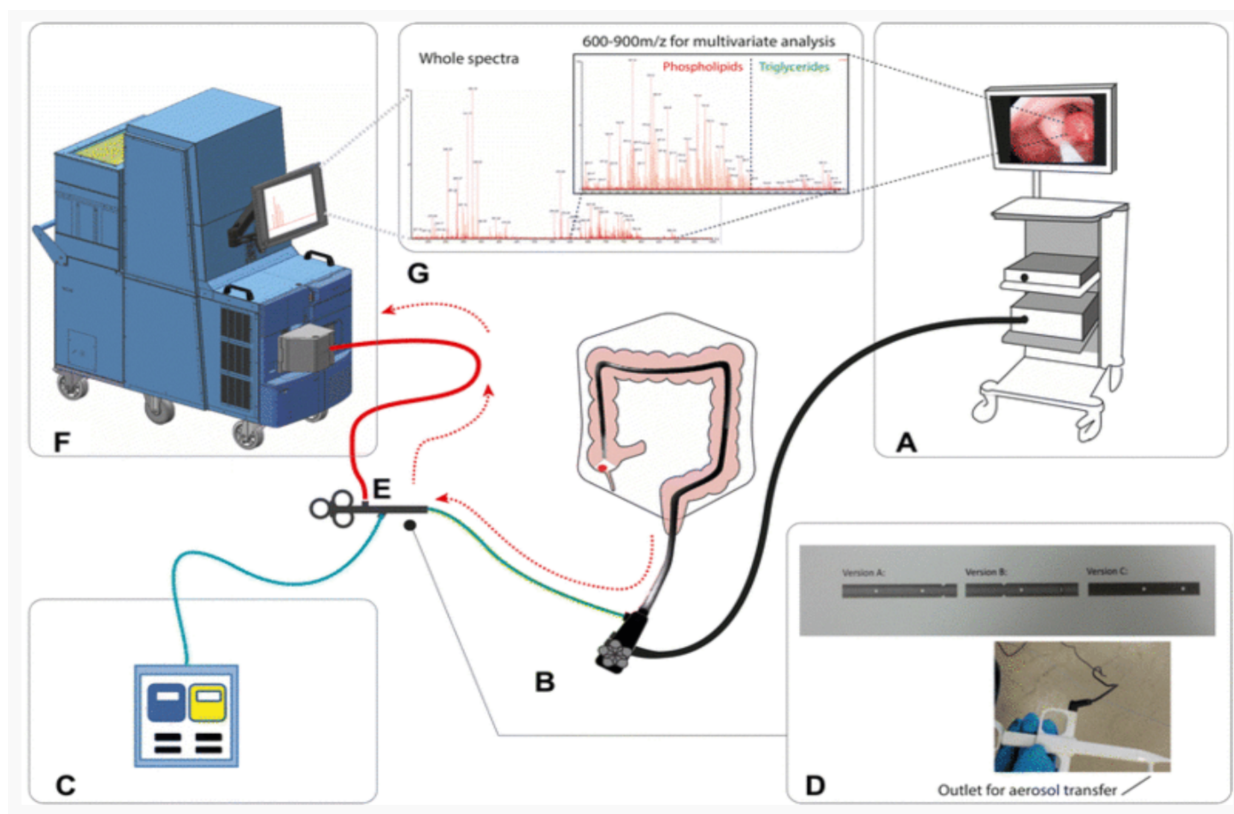


Figure 9. in vivo endoscopy using the iKnife.

A: typical stack. B: commercially available endoscope. C: standard electrosurgical generator is deployed during 'hot' endoscopic resection. D: Modified snare with fenestrated distal sheath for improved efficiency of smoke aspiration. F: a mobile mass spectrometer, with a self contained spectral database sited in the endoscopy suite. G: An adenoma snared during an in vivo assessment of the REIMS platform with representative anatomically discrete lipidomic data captured during hot snare deployment. Taken from (Alexander et al., 2017).

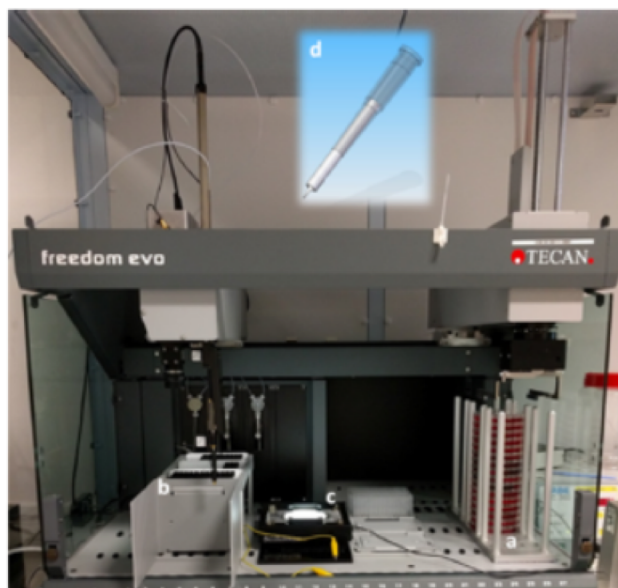
2.5.3.5. Microbiology

Before the widespread introduction of mass spectrometry platforms, such as matrix-assisted laser desorption ionisation time of flight (MALDI-ToF), into clinical microbiology laboratories, the majority of identifications relied upon extensive biochemical characterisations (Seng et al., 2010). These included sugar utilisation and fermentation; but were time consuming, variable, and required expertise to perform. The implementation of mass spectrometry platforms has allowed for rapid microbial identification; improving patient care through reduced diagnosis times, and targeted clinical interventions. MALDI-ToF based platforms are currently the primary systems used in clinical diagnostic laboratories. Microbial identifications using this method rely on the characterisation of ionised proteins within the mass range of 2–20 kDa; with over 50% of ribosomal origin (Croxatto et al., 2012). Although, the introduction of MALDI-ToF has reduced the time to identification of microbial colonies, it still requires a user to add a matrix to assist in ionisation. Furthermore, in some instances, such as for yeasts, additional extraction steps are required for accurate species level identification (Cassagne et al., 2013). MALDI-ToF based

systems can also have difficulty in the species-level classification of some taxonomic groupings, such as *Salmonella* sp., and in sub-species analysis.

The application of REIMS to the analysis of microorganisms naturally follows its development as a diagnostic tool in cancer margin assessment. This differentiation relies primarily on the analysis of phospholipids and other complex lipids, which have, prior to the use of ribosomal proteins in MALDI-TOF, been used in the speciation of both bacteria and yeasts (Abel et al., 1963). Similar to iKnife applications employing REIMS, early work on the analysis of microorganisms employed modified surgical tools. The handheld monopolar iKnife handpiece was used, but due to the high electric current density poor signal quality was achieved, particularly at higher mass ranges. In contrast, a handheld bipolar forceps set-up, whereby microbial biomass is heated between two probe tips after removal from agar culture plate, achieved a more evenly distributed electrical current and thus heating effectiveness (Strittmatter et al., 2013) (Strittmatter et al., 2014). This approach proved highly effective in microbial speciation of 28 clinically important bacterial and yeast species. Although an effective approach for use in research and technical development, the handheld bipolar forceps used in REIMS analysis of microorganisms are likely unsuitable for clinical diagnostics labs due to the high user involvement, requirement to clean probe surfaces between analyses, and limited analytical throughput.

To increase analytical throughput and to reduce operating costs, diagnostic microbiology laboratories have undergone rapid automation due to increased financial pressures, centralisation of diagnostic laboratories, and substantially increased numbers of diagnostic specimens (Bourbeau and Ledeboer, 2013). Due to the increasingly high-throughput nature of clinical microbiology laboratories, to be clinically utilised REIMS is required to fit into these automated workflows. To this end, an automated high-throughput REIMS platform, incorporating a Pickolo visualisation platform and sample handling capability into a TECAN Freedom EVO robot, Photograph 1. This platform minimises user related errors and thus makes it highly applicable to the workflow of modern clinical microbiology laboratories; possessing the capability of analysing thousands of microbial colonies without any preparative steps in a 24 hour period (Bolt et al., 2016) (Bolt et al., 2016, Cameron et al., 2016).



Photograph 1. Customized TECAN Freedom EVO automated sample handling machine and sampling probe.

A customized TECAN Freedom EVO automated sample handling machine is contained within a class 2 microbiological safety cabinet. Before use, agar culture plates are placed in a plate rack (a), and sampling probes are loaded into a 96 rack (b). Plates are then moved from the plate rack to the Pickolo visualization platform (c), which also acts as the return electrode, where the user or the automated system can select colonies for analysis. A technical drawing of the stainless steel sampling probe is given in (d).

2.5.4. Conclusions

The clinical applications of the iKnife have shown that ambient spectroscopy is flexible and highly amenable to challenging clinical environments. In an era where precision surgery is developing fast - meaning the right operation at the right time and in the right patient with the best outcome - the powerful ability of the iKnife to provide real-time phenotypic data on tumours makes it an ideal platform for clinical use.

The rich cancer lipidome data collected during clinical use leading to the identification of phenotypic classes makes the iKnife an ideal research tool for studying the cancer lipidome and for novel biomarker discovery.

Further areas of medicine outside of oncology have also benefited from the ability of on-demand clinical mass spectrometry at the point of care for answering complex clinical questions. In particular, microbiology could potentially be revolutionised by the ability, in some cases, to immediately classify micro-organisms without the need for time consuming culture.

The next steps for the iKnife will involve further clinical trials of new versions of equipment in larger patient populations, collecting more data to produce larger, more

accurate databases for phonemic characterisation. This data may be able to answer some key questions regarding the *in vivo* heterogeneity of tumours as well as the clinical correlates of novel molecular signatures that could lead to a revolution in the way diseases are profiled and treated.

2.6. Raman Spectroscopy: an introduction

This section aims to introduce Raman spectroscopy and to outline its development from an observed physical phenomenon to a means for the rapid, preparation-free and non destructive molecular analysis of biological tissue during surgery.

2.6.1. Innovation after innovation

It was on the 28th of February, 1928 that CV Raman would discover the radiation effect which bears his name: "A new radiation", Indian J. Phys., 2 (1928) 387. His observation was that a small amount of monochromatic light, when passing through a transparent liquid, was scattered by the liquid at a different wavelength. The effect was weak (fewer than 1 in a million photons are scattered in this way) but was observed for over 60 liquids and was thus a universal phenomenon.

Since this discovery a number of innovations have taken place to make measuring the Raman effect faster and easier. The innovations and discoveries which have led to present day use of Raman spectroscopy are detailed in Figure 10. These steps include the discovery of lasers and the Charge Coupled Device (CCD) for which Nobel Prizes were awarded. The most recent addition to this is the widespread use of machine learning which has enabled classification models to be built using the highly complex Raman spectra obtained from biological tissue samples. This is an important precursor to the use of Raman spectroscopy in surgery.



2018	ExoMars Rover, taking Raman spectroscopy to Mars
2010	Video-rate <i>in vivo</i> SRS Raman imaging by Sunny Xie
2010	SESORS reported by Graham, Faulds, Stone and Matousek
2009	Boyle and Smith awarded Nobel Prize in Physics for CCD
2005	SORS invented by Pavel Matousek
2000	TERS developed by Volker Deckert
1997	Single molecule SERS reported by Katrin Kneipp
1990s	Significant increase in commercial instrumentation
1983	Enhancement of 10^{14} reported by Richard van Duyne
1982	Use of Ag and Au colloids for SERS by Lee and Meisel
1980s	CCD detectors used for Raman
1980s	Fibre optics coupled with Raman
1974	Fleischman and colleagues observe surface enhanced Raman scattering (SERS)
1970s	Raman coupled with microscopy
1969	Invention of charge-coupled device (Willard Boyle and George Smith)
1964	Townes, Basov and Prochorov awarded Nobel Prize in Physics for invention of the laser
1960s	Stimulated Raman and coherent anti-Stokes Raman (CARS) invented
1960s	Lasers used as light source for Raman spectrometers
1957-59	Light amplification by stimulated emission of radiation constructed
1953	First commercial Raman spectrometer
1950	Observation of resonance Raman (RR) spectra
1940s	Invention of Raman grating spectrometer
1930s	Invention of the monochromator
1930	Sir Chandrasekhara Venkata Raman awarded Nobel Prize in Physics
1928	C.V. Raman and K.S. Krishnan: discovery of Raman effect experiments on 28 Feb
1928	Landsberg and Mandelstam: independent observation of inelastic light scattering
1927	Compton awarded Nobel Prize in Physics
1923	Arthur Compton (effect): discovered inelastic X-rays & γ -rays scattering in matter
1923	Adolf Smekal's prediction of the Raman effect
1922	CV Raman's monograph published on the Molecular Diffraction of Light
1921	CV Raman's experiments on the colour of the sea, showing that the colour was due to molecular diffraction
1917	Albert Einstein first theorizes about stimulated emission (On the Quantum Theory of Radiation)
1871-1899	Lord Rayleigh (John Strutt) further refines his theory of scattering in a series of papers
1871	Rayleigh scattering first described

Year

Figure 10. A timeline of the key discoveries leading to present day uses of Raman spectroscopy.

Dates in red indicate Nobel Prizes. Adapted from Ellis et al. (Ellis et al., 2013)

2.6.2. The chemistry of light scattering

When light strikes an object, it interacts with the molecules in that object and can be reflected (the angle of incidence being equal to the angle of reflection) or scattered (the angle of incident light being different to that of the scattered light). The majority of scattered light involves no exchange of energy between the incident photon and the molecule it is striking and the scattered photon is the same wavelength of the incident photon (elastic scattering) whereas a small number of photons exchange energy with the molecule they are striking and are scattered at a different wavelength (inelastically scattered light). This inelastically scattered light is observed in the Raman Effect. This effect is unique in currently used optical methods in that it can be analysed to produce molecular information unique to the chemical bonds being studied and their surrounding environment.

To be Raman active a molecule must have anisotropic polarisability (polarisability which is different in different directions). Of note is the fact that water does not exhibit anisotropic polarisability and so is not Raman active. Molecule polarisability can be thought of as an interaction between the electron clouds of a chemical bond and the surrounding environment. Chemical bonds in a simple diatomic molecule can be thought of as a simple harmonic oscillator, with the formula of square root K over μ reacting to the spring constant and mass of the molecule. The spring constant here is the bond strength which is an electromagnetic interaction between the molecules and the local environment. An example of the importance of the interaction with the

local environment is the Raman spectrum of acetone when in the gas phase and when in a water mixture. The molecule is the same, except in water there is a drop in frequency and broadened peaks as the chemical bond is slightly weaker in the presence of water.

The basis of the photon-chemical bond interaction of Raman provides a unique molecular window into materials being analysed. Raman is used as part of a number of different systems depending on the aim of the chemical analysis being performed.

2.6.3. Modern Raman systems and applications

Current modern day systems consist of the following hardware components, as illustrated in Figure 11:

A laser light source to provide powerful monochromatic light

Optical filters to help detect wavelength change of scattered light- these have a specific optical coating

A Detector- often a cooled CCD device

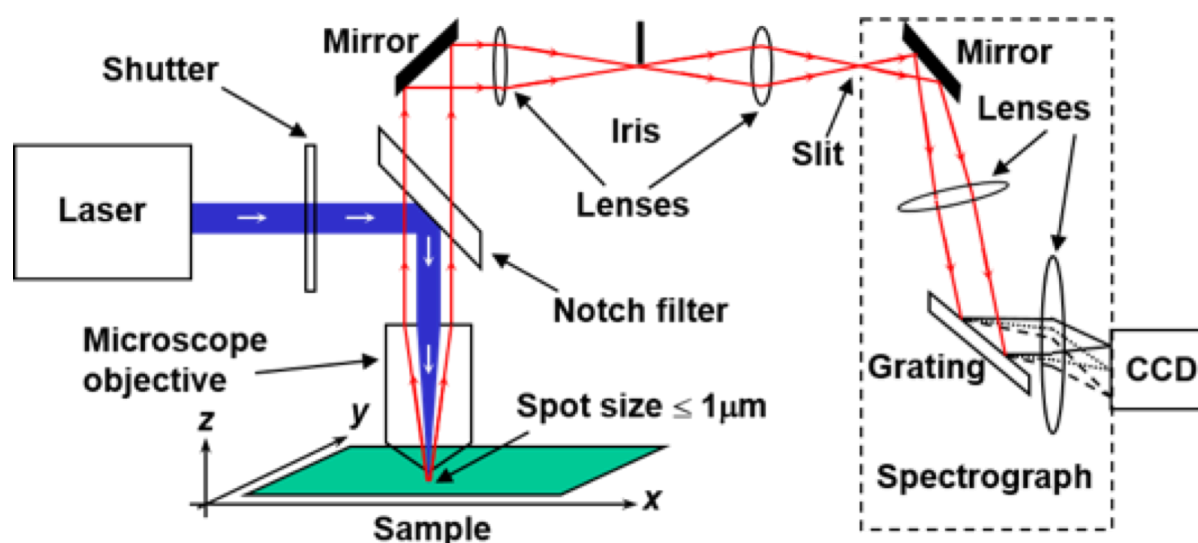


Figure 11. Raman spectrometer schematic.

A narrow line-width laser source at a particular wavelength, and optimized to the material under investigation, is focused on the sample. Scattered light is collected by the microscope lens and passes through a notch or edge filter which eliminates light of the original wavelength, leaving only inelastically. Reproduced from Short et al, 2006, with permission from Springer.

The majority of Raman systems in use are laboratory benchtop systems which provide a controlled environment for Raman readings to be obtained. Raman microscopes allow the sample to be placed on a stage for analysis and Raman imaging to be performed. Current examples of the successful use of Raman spectroscopy include silicon detection and analysis - looking at crystallinity - pharmaceuticals and explosives detection and medical uses. Raman is used in the pharmaceutical and explosives detection industry for non-quantitative, quantitative

and semi-quantitative analysis of drugs and other chemicals. It can be used to assess purity of drug preparations. It is also used in the food industry to for food quality and safety assessment without opening packaging or destroying consumable merchandise. The speed of Raman enables its use in real time integrated into the production line screening of products being prepared.

2.6.4. Raman Spectroscopy in biological systems

Raman offers a non-invasive, non-destructive, label free analysis method which is unaffected by water and hence lends itself well to biological applications. Raman spectroscopy has been proven in multiple studies in mice and humans to accurately diagnose cancerous changes in tissue specimens taken from diseased brains and analysed using large laboratory based Raman microscopes. Glioblastomas implanted in rat brains (Beljebbar et al., 2010) and mice brains (Ji et al., 2013) are easily detectable with high sensitivity and specificity (>99%) using Raman Spectroscopy. This paved the way for human samples to be analysed. Leslie et al. (Leslie et al., 2012) proved that human brain tumours have unique Raman Spectra which can be used for diagnosis in tissue samples taken from surgery and analysed afterwards in the lab. Gajjar and colleagues also demonstrated that Raman spectroscopy could accurately diagnose human brain tumours in a sample of 52 paraffin embedded blocks taken from patients during surgery (Gajjar et al., 2012a). These studies show that the high accuracy and precision of Raman Spectroscopy matches or even exceeds that of routine histology which is the current gold standard. The stage is set for a Raman spectroscopic instrument which could be used for histological evaluation in situ rather than in the lab which would give the surgeon immediate real-time information during surgery of the nature of the tumour and its margin. Knowing the diagnosis of the tumour immediately would also revolutionise surgery as the surgeon can make an instant decision not to resect all of a radiosensitive tumour if it is in a problematic area of the brain which can easily be targeted by post operative radiotherapy.

2.6.5. The need for new Raman technology

We now have the ability to introduce, for the first time ever, a Raman probe small enough to be used within the brain tumour resection cavity during surgery to help instantly diagnose tumour tissue and the tumour/brain border. This builds on all of the basic science and animal research which has proven that Raman spectroscopy can give highly accurate diagnostic data, and recent reports on human tissue samples analysed post operatively which also agree on this.

The previous few sections have introduced Raman spectroscopy and REIMS. The next section will describe the hypothesis tested using both of these technologies in this thesis.

2.7. The hypothesis

My main hypothesis to be tested in the work in this thesis is that molecular analysis at the time of surgery in real time is possible using Raman spectroscopy and Rapid Evaporative Ionization Mass Spectrometry (REIMS).

My secondary hypothesis is that real time chemometric analysis can reveal characteristics of the tissue being operated on which are relevant to the surgeon and treating oncology team.

The underlying basis for both of these hypotheses is that the analytical methods being proposed are unique in that they both collect molecular information almost instantly. As cancer is by its nature a phenotype emerging from an interaction of genes and the environment the techniques being developed in this thesis should be able to collect useful biological information from the tissue being analysed at the time of surgery.

2.8. Thesis objectives and outline

The aim of this thesis is to show how surgical research should move from the laboratory to benefit the patient on the operating table as a continuous body of work over 4 years. Specifically, my objectives are:

1. To develop 2 new chemometric technologies for 'First in Human' use
2. To develop new methodologies for the real-time molecular profiling of brain tumours during surgery
3. To use this new paradigm for the *in vivo* modelling of brain cancer to answer key clinical questions: a platform for precision medicine

My secondary objectives are:

1. To build experience in leading 2 new surgical trials in neuro-oncology
2. To lead by example through innovation in neurosurgery
3. To raise awareness and stimulate cross-disciplinary research in brain cancer

The introductory chapters have introduced the reader to the field of neuro-oncology and the surgical management of this disease and finished with the key clinical questions which exist in this field for the treating surgeon.

Next this thesis will contain a section on a systematic review of *in vivo* tumour detection technology in brain surgery to establish the benchmark against which Raman spectroscopy and REIMS must be measured. The thesis will then treat both *in vivo* technologies being developed in 2 separate sections.

The Raman spectroscopy section will detail the development of this technology for use during surgery. Important aspects of system design and development, operating theatre setup, calibration and first use during surgery will be explored. Data analysis approaches using machine learning will be discussed and the data obtained from first surgical use of the system will be presented.

The following section will deal with Rapid Evaporative Ionization Mass Spectrometry (also called the iKnife). The development of a brain surgery iKnife system will be described together with the development of system setup, calibration and standardisation procedures. Data analysis methodology for mass spectral data will be discussed and developed. The results of the first surgical use of the system in brain surgery will be presented.

The final section will deal with the discussion of the data and issues raised by the use of these 2 novel technologies in brain surgery. The success of use of the techniques will be assessed as well as lessons learned regarding diagnostic studies in this field. This will be the end of a journey into delivering 2 novel technologies and disciplines of science from the laboratory to the operating theatre in an effort to move neurosurgery into a new era of precision medicine.

3. Systematic review on intraoperative tissue detection

3.1.1. Introduction

A number of technologies are emerging with the potential for intraoperative diagnosis in brain tumour surgery which need to be assessed for effectiveness. There are no previous systemic reviews or meta analyses in this area to guide benchmarking of these new technologies. There is currently a focus on improving EOR in brain tumour surgery in an effort to improve survival in glioma. During surgery tumour tissue is diagnosed using frozen section, but the use of this technique is not universal: due to time, cost and resource implications of histopathological analysis. The gold standard for diagnosis is permanent section histopathology.

Successful translation of emerging techniques to clinical practice will rely upon achieving diagnostic accuracy at least comparable to existing methods, while also providing benefits such as rapid diagnosis and improved cost-effectiveness. The aim of this study is to identify and evaluate the accuracy of current technologies which have been used for tumour tissue analysis during surgery for brain tumours as a baseline for new emerging technologies. Through a review of existing technologies, the challenges in brain tumour detection can also be outlined to help guide future attempts at *in vivo* tumour detection.

3.1.2. Methods

This was a systematic review and meta-analysis of intraoperative brain tumour diagnosis techniques compared to the diagnostic accuracy of permanent section histopathology. The primary outcome measure was diagnostic accuracy. Secondary outcome measures included the turnaround time for the result.

3.1.2.1. Search string

A systematic review was conducted using PRISMA guidelines. The last search was conducted in March 2017 on Pubmed, Cochrane Library database, EMBASE and Scopus. Where possible, MESH terms were used and all-text searches were performed for the following string:

- 1 exp Glioma/ (187118)
- 2 Glioma\$.mp. (138052)
- 3 or/1-2 (216387)
- 4 exp "Sensitivity and Specificity"/ (804834)

5 "sensitivity and specificity"/ (610292)
 6 "sensitivity and specificity".mp. (722039)
 7 (sensitivity or specificity).mp. [mp=ti, ab, hw, tn, ot, dm, mf, dv, kw, fx, nm, kf, px, rx, ui, sy]
 (3099275)
 8 or/4-7 (3262162)
 9 3 and 8 (15494)
 10 Intraoperative Period/ (46935)
 11 (intraoperative\$ or intra-operative\$).mp. [mp=ti, ab, hw, tn, ot, dm, mf, dv, kw, fx, nm, kf, px, rx,
 ui, sy] (352413)
 12 (during adj3 surg\$).mp. [mp=ti, ab, hw, tn, ot, dm, mf, dv, kw, fx, nm, kf, px, rx, ui, sy] (176544)
 13 or/10-12 (493157)
 14 9 and 13 (606)
 15 limit 14 to english language (587)
 16 Comment/ or Letter/ or Editorial/ (3135209)
 17 (comment or letter or editorial).pt. (3172422)
 18 case reports.pt. (1868409)
 19 or/16-18 (4880006)
 20 15 not 19 (559)
 21 remove duplicates from 20 (414)
 22 conference.pt. (3664428)
 23 meta-analysis.pt. (85374)
 24 review.pt. (4688390)
 25 or/22-24 (8382537)
 26 21 not 25 (301)

Further studies were identified through bibliographic cross referencing, Google Scholar searches and personal knowledge of the field until no further studies meeting the inclusion criteria could be identified. All relevant citations meeting the inclusion criteria were selected for quality assessment through full text review.

3.1.2.2. Inclusion criteria

Studies written in English that contained data on glioma characterisation or diagnosis during brain tumour surgery were eligible. Only those studies containing data on sensitivity and specificity compared to definitive permanent histopathology were included. Adults and children were included as study participants. The last search was performed in March 2017.

3.1.2.3. Exclusion criteria

Abstracts, conference articles, opinions, case studies, reviews and meta-analyses were not to be considered. Studies not performed in humans or in a clinical environment were to be excluded. Studies failing to report sensitivity and specificity compared to permanent section histopathology were to be excluded.

3.1.2.4. Study quality assessment

Full text documents of all shortlisted studies were collected and study methodology was assessed for bias using the QUADAS-2 validated scoring system designed

specifically for primary diagnostic studies (Whiting et al., 2011). QUADAS-2 scores for 7 signalling questions were assessed on a 14 point system, with 1 point given for low risk to bias and 2 points given for high or unclear risk to bias. The signalling questions were used after BV, AW and HK reached agreement on scores after applying the questions to a test sample of studies with the lowest possible score of 7/14 indicating the lowest possible bias.

3.1.2.5. Data collection

Each of the selected studies were included in a table comparing study characteristics (Table 4). Data collected included (1) the first author; (2) year of publication; (3) study design; (4) technology type (5) *in vivo* or *ex vivo* technique (6) destructive or non-destructive data collection (7) result turnaround time.

The 3 investigators extracted data independently and diagnostic accuracy figures were calculated to 1 decimal point. Data was collected in Excel spreadsheet format with drop down boxes used where possible to standardise data entry.

3.1.3. Results

3.1.4. Search strategy and selection criteria

The initial search revealed 301 citations. Searching the Cochrane review yielded 2 systematic reviews on related subjects (image detection technologies for improved EOR (Jenkinson et al., 2018) and resection of brain tumours (Barone et al., 2014)). 20 duplicates were excluded, leaving 284 citations. Abstracts from the 284 citations were reviewed. 193 abstracts did not meet the inclusion criteria, specifically because

1. Sensitivity and specificity data was not included
2. The paper was a literature review or a systemic review
3. The study was done on animals, with no clinical data on humans
4. Intraoperative techniques were not assessed

77 abstracts fulfilled the criteria for this review- full texts were obtained of all of these papers. Upon reading full texts 5 papers did not meet the inclusion criteria, hence 72 studies were included in the meta analysis.

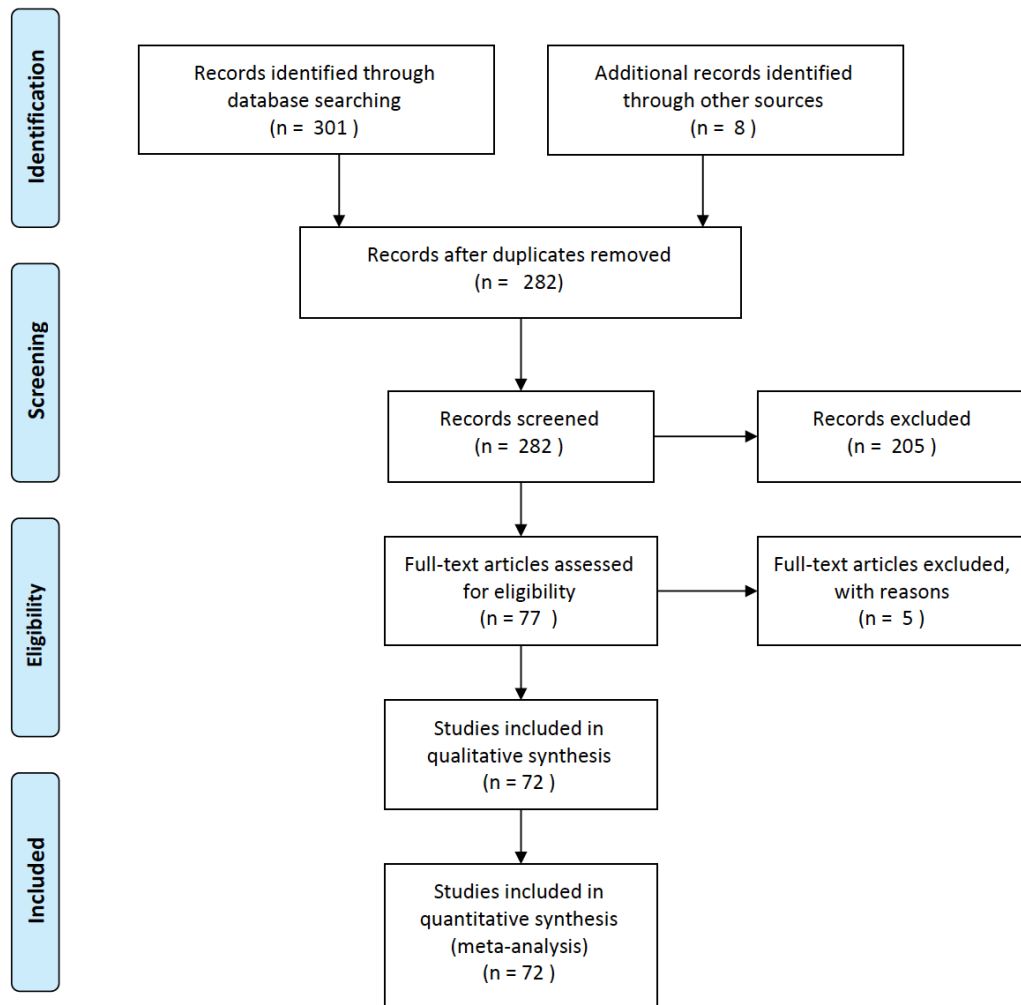


Figure 12. PRISMA 2009 flow diagram showing the study selection process for meta-analysis.

3.1.5. Study demographics

Study demographics

72 studies were identified including data from a combined total of 11593 patients and 15072 tissue biopsies. In terms of study design, 48 prospective studies and 24 retrospective studies were included. Patients were selected as a consecutive series in 20 studies, randomly in 1 study and in a selected fashion in 39 studies. The studies were published over a 20 year period from 1998 to 2018. 24 studies contained data identifying patient age, 41 were of *in vivo* methodology and 36 were *ex vivo* methodology with 2 studies having both (shown graphically in Figure 13). A number of types of technology were used in the studies included in this meta-analysis. The broad classes of technology are illustrated in Figure 14 and can be divided into imaging (20 studies), histopathology (24 studies), optical technology (32 studies) and mass spectrometry (3 studies). 5 studies used a combination of technologies. A breakdown of the 22 sub-types of technology is shown in Table 3.

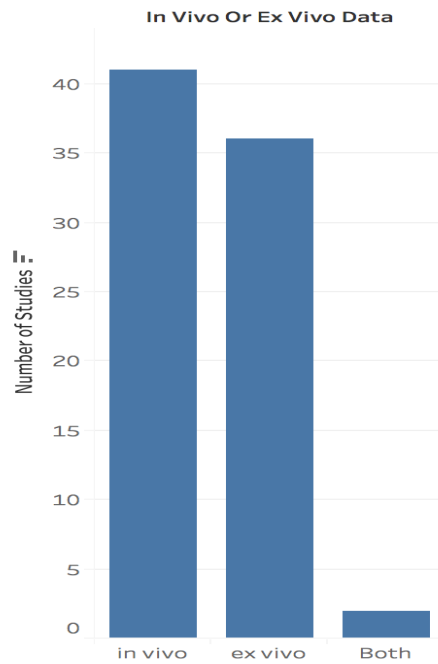


Figure 13. Bar chart showing number of *in vivo* and *ex vivo* studies used for intraoperative brain tissue detection.

Histopathology

- Frozen section
- Smear cytology
- Squash cytology
- Touch cytology
- Flow cytometry
- Rapid immunohistochemistry

Optical

- Optical coherence tomography
- Light scattering microscopy
- Spectroscopic probes
- 5-ALA fluorescence
- Fluorescein fluorescence
- Raman spectroscopy
- Diffuse reflectance microscopy
- Confocal laser endomicroscopy
- ICG second window
- Infrared spectroscopy

Imaging

- Intraoperative MRI
- Neuronavigation on pre-operative MRI
- Magnetic resonance spectroscopy
- Intraoperative ultrasound
- PET imaging

Mass spectrometry

- DESI

Table 3. The broad groups of technology and subtypes of technology used for intraoperative brain tissue detection.

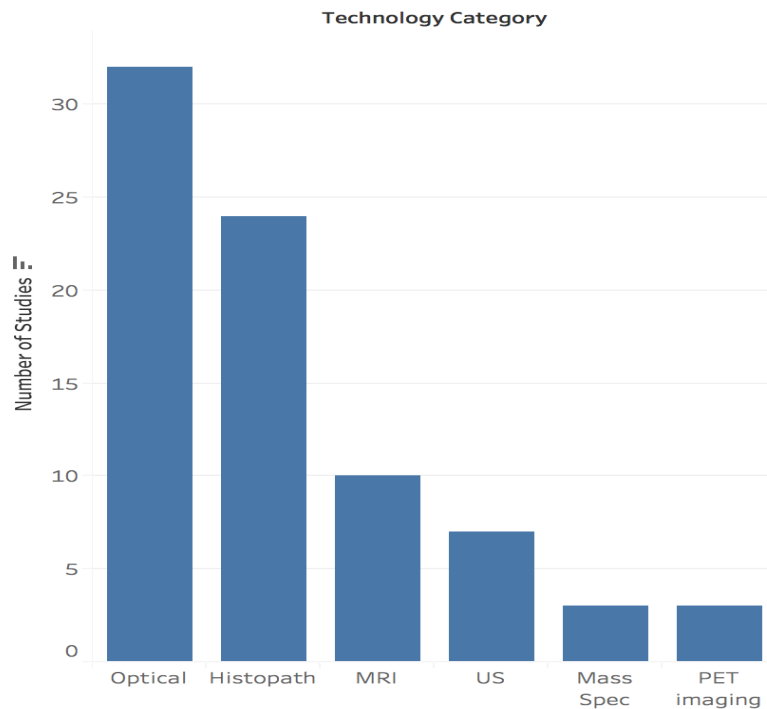


Figure 14. Distribution of studies with respect to technology category for intraoperative brain tissue detection.

It can be seen that optical technologies are the most investigated modality for intraoperative brain tissue detection, followed by histopathological techniques and MRI. Within the optical group, 5-ALA was the most investigated technique. In histopathology the most investigated technique was frozen section and in MRI studies intraoperative MRI was the most investigated subtype of this technology.

3.1.6. Main results

The main technology groups will be described together with the diagnostic accuracy findings. Table 4 contains the main dataset.

Frozen section

8 studies reported on the accuracy of frozen section. These were all *ex vivo* retrospective studies on large numbers of patients (1634) in low and high grade gliomas. The average sensitivity was 89% with 98% average specificity with a lowest reported sensitivity of 64% and specificity of 96%.

Cytology

9 studies containing data from 4172 patients reported on the use of cytology. All apart from one study (Khamechian et al., 2012) were *ex vivo* retrospective design. The average sensitivity was 94.2% and specificity was 89.6%, with a lowest reported sensitivity of 84% and specificity of 88%.

Rapid immunohistochemistry

1 study enrolled 183 patients as part of a prospective trial to investigate rapid immunohistochemistry. Samples were processed in the laboratory with a sensitivity

and specificity of 90%.

Flow cytometry

2 studies investigated flow cytometry for diagnosis of brain tumours during surgery for 372 patients. The studies were *ex vivo* prospective designs with the technique applied intraoperatively. An average sensitivity and specificity of 88% was reached using this method.

MRI

Intraoperative MRI was used by 5 studies involving a total of 101 patients. These studies were prospective, based on a consecutive series of patients and the data was collected *in vivo* with an MRI scanner in the operating room. Biopsies were taken under neuronavigation guidance using the updated MRI imaging dataset to take into account brain shift. The averaged sensitivity using a range of MRI modalities was 71% and specificity reached 81%.

Ultrasound

6 studies investigated the use of ultrasound during surgery for the identification of tumour tissue. These *in vivo* prospective studies recruited a combined total of 66 patients. Only 1 study enrolled consecutive patients with all other studies selecting patients for use with the ultrasound systems, potentially introducing bias. QUADAS scores were however low indicating no other sources of bias and high grade and low grade gliomas were included. The range of technology included navigated and non-navigated ultrasound and reached an average sensitivity of 67.8% and specificity of 76.8%.

PET Imaging

3 studies used PET imaging to help target tumour tissue during surgery. Methionine PET and F-PET were used in 30 patients as part of prospective studies in gliomas. An average sensitivity of 89% and specificity of 82.3% was achieved.

Optical technologies

The diagnostic accuracy of 5ALA was investigated in 9 studies on its own and in 3 studies with a spectroscopic probe. 300 patients were enrolled in total in these *in vivo* prospective trials. 5ALA alone reached a sensitivity of 86% and specificity of 85%. With spectroscopic probes it reached a sensitivity of 86% and a specificity of 73%.

Fluorescein was investigated in 2 prospective studies in GBM where data was collected *in vivo*. These studies involved a total of only 44 patients and achieved an average sensitivity of 86.5% specificity of 94.5%.

Confocal microscopy and endoscopy were used in 4 studies. These studies were prospective and used *in vivo* and *ex vivo* data collection. An averaged sensitivity of 88.8% and specificity of 93% was obtained in these studies.

Reflectance and light scattering spectroscopy were used in small groups of patients in 2 studies as part of prospective trials on *ex vivo* samples. An averaged sensitivity of 90% and 91.5% was achieved for all gliomas.

The diagnostic accuracy of Raman spectroscopy was investigated using *ex vivo* Raman microscopy in 2 studies and an *in vivo* handheld Raman probe in one study. At least 63 patients were recruited in these prospective studies which included patients with gliomas and metastases. An average sensitivity of 96.5% and specificity of 95.5% was reached.

Mass spectrometry

Mass spectrometry was investigated by 3 studies on *ex vivo* tissue analysed by DESI. 85 patients were enrolled in these studies which looked at low and high grade gliomas. The average sensitivity was high at 96.1% and specificity was 98.6%.

Technology Category	Study Name	Technology Method	In vivo or ex vivo data	Destructive/ Non destructive data collection from tissue samples	Number of Patients	Number of biopsies/Study Design	Patient selection	Disease type	Mean Age	QUADAS-2 Score	Turnaround time for result	Index Test	Tumour Sensitivity	Tumour Specificity
Combination: S-AIA and MRI	Hauser 2015	S-AIA resection then MRI	in vivo	Non-Destructive	14	117 Prospective	Selected	HGG			7 unclear	Histopathology	0.81	0.43
Combination: S-AIA and MRI	Gessler 2015	S-AIA resection then MRI	in vivo	Non-Destructive							8	Histopathology	0.75	1
Combination: S-AIA and MRI	Colubrier 2014	S-AIA resection then MRI	in vivo	Non-Destructive							not reported	Histopathology	0.75	0.80
Combination: S-AIA and NeuroNavigation	Wang 2012	S-AIA and NeuroNavigation	in vivo	Non-Destructive								Histopathology	0.98	0.46
Combination: PET and S-AIA	Ewelt 2011	FET-PET + S-AIA	in vivo	Non-Destructive	30	39 Prospective	Consecutive	HGG	42.1	7	N/A	Histopathology	0.70	0.92
Histopathology	Patric 1999	Touch and smear cytology	ex vivo	Destructive	595	595 Retrospective	Consecutive	CNS tumours			30 mins	Histopathology	0.96	0.75
Histopathology	Chand 2016	Crash Preparation and Frozen Section	ex vivo	Destructive	80	80 Prospective	Selected	CNS tumours			70 mins	Histopathology	0.95	1
Histopathology	Khamachian 2012	Touch cytology	ex vivo	Destructive	139	139 Retrospective	Selected	CNS tumours	57			Histopathology	0.84	0.88
Histopathology	Ishikawa 2014	Frozen section	ex vivo	Destructive	205	205 Retrospective	Selected	Gliomas			7 20 mins	Histopathology	0.91	0.98
Histopathology	Hamasaki 2018	Touch and smear cytology Diff-Quik stain	ex vivo	Destructive	400	400 Retrospective	Consecutive	Brain tumours			7 30 mins	Histopathology	0.979	1
Histopathology	Fauzi 2015	Optical cytology analysis	ex vivo	Destructive	53	53 Retrospective	Selected	Brain tumours			8 30 mins	Histopathology	0.88	0.85
Histopathology	Kishinami 2012	Squash smear cytology	ex vivo	Destructive	865	865 Retrospective	Consecutive	CNS Tumours	32.2		8 unclear	Histopathology	0.987	0.95
Histopathology	Briggs-Torres 2001	Smear cytology	ex vivo	Destructive	650	650 Retrospective	Consecutive	CNS Tumours			8 unclear	Histopathology	0.979	0.95
Histopathology	Tanino 2015	Rapid immunohistochemistry	ex vivo	Destructive	183	183 Prospective	Consecutive	Brain tumours			9	Histopathology	0.9	0.81
Histopathology	Shoyama 2013	Flow Cytometry	ex vivo	Destructive	139	139 Retrospective	Selected	CNS Tumours			10 20 mins	Histopathology	0.86	0.91
Histopathology	Tette 2014	Frozen section	ex vivo	Destructive	578	578 Retrospective	Consecutive	HGG	57		8	Histopathology	0.791	0.955
Histopathology	Tette 2014	Frozen section	ex vivo	Destructive	578	578 Retrospective	Consecutive	HGG	57		8	Histopathology	0.644	0.967
Histopathology	Bruckner 2016	Frozen section	ex vivo	Destructive				HGG				Histopathology	0.90	1.00
Histopathology	Bruckner 2016	Frozen section	ex vivo	Destructive				HGG				Histopathology	1.00	1.00
Histopathology	Kanyama 2018	Flow Cytometry	ex vivo	Destructive	250	250 Retrospective	Selected	Brain Tumour N/A			7 <10 minutes	Histopathology	0.89	0.94
Histopathology	Kanyama 2018	Squash Cytology	in vivo	Destructive	139	139 Retrospective	Not reported	Gliomas	57		7 N/A	Histopathology	0.84	0.88
Histopathology	Shoyama 2013	Flow Cytometry	ex vivo	Destructive	81	128 Prospective	Selected	Gliomas	men (mean)		8 15 minutes	Histopathology	0.88	0.88
Histopathology	Shoyama 2013	Frozen section	ex vivo	Destructive							20 minutes	Histopathology	1.00	1.00
Histopathology	Altanbulut 2016	Microfluidics	ex vivo	Destructive	47	47 Prospective	Selected	Gliomas	45.70 ± 12.6		9 30 minutes	Histopathology	0.88	0.64
Histopathology	Martinsen 2018	Frozen section	ex vivo	Destructive		126						Histopathology	0.95	0.96
Histopathology	Khodami 2015	Frozen section	ex vivo	Destructive	273	273 Retrospective	Consecutive	Gliomas	35.1		7 N/A	Histopathology	0.91	0.99
Histopathology	Endogan 2014	Squash Cytology	ex vivo	Destructive	545	545 Retrospective	Consecutive	Gliomas	N/A		9 not reported	Histopathology	0.99	0.92
Histopathology	Karl Bassler, M.D.	Smear Cytology	ex vivo	Destructive	4172	4172 Retrospective	Not reported	Gliomas	N/A		8 not reported	Histopathology	0.90	not reported
Histopathology	Briggs-Torres 2001	Squash Cytology	ex vivo	Destructive	650	650 Retrospective	Consecutive	CNS Tumours N/A			9 N/A	Histopathology	0.98	0.95
Optical	Samdhi 2013	multi-modal confocal imaging	ex vivo	Non-destructive	37	37 Retrospective	Selected	Brain tumours				Histopathology	0.88	1
Mass Spectrometry	Jarmusch 2016	DESI	ex vivo	Destructive	49	49 Retrospective	Selected	Gliomas			30 60 minutes	Histopathology	0.943	0.985
Mass Spectrometry	Eberlin 2011	DESI	ex vivo	Destructive	36	242 Retrospective	Selected	Gliomas			Hours	Histopathology	0.979	7
Mass Spectrometry	Eberlin 2012	DESI	ex vivo	Destructive	5	26 Prospective	Selected	Brain tumours			1 hour	Histopathology	0.997	7
MRI	Hu 2009	PCV: SWI Contrast Enhanced perfusion MRI	in vivo	Non-Destructive	13	13 Prospective	Selected	HGG			11 60 mins	Histopathology	0.917	100
MRI	Hesselmann 2017	Intraoperative MRI	in vivo	Non-Destructive	19	19 Prospective	Selected	HGG			9 mins	Histopathology	0.95	0.695
MRI	Colubrier 2017	Intraoperative MRI	in vivo	Non-Destructive	68	68 Prospective	Selected	HGG	59		8	Histopathology	0.83	0.67
MRI	Colubrier 2015a	Intraoperative MRI	in vivo	Non-Destructive	33	99 Prospective	Selected	HGG	60			Histopathology	0.93	0.93
MRI	Colubrier 2015b	Intraoperative MRI	in vivo	Non-Destructive								Histopathology	0.55	0.74
MRI	Virtikuntti 2013	Magnetic Resonance Spectroscopy	ex vivo	Non-Destructive	58	58 Prospective	selected	Gliomas	57		9	Histopathology	0.71	0.76
MRI	Pancani 2012	NeuroNavigation	in vivo	Non-Destructive								Histopathology	0.56	0.57
MRI	Pamir 2013	ioMRI	in vivo	Non-Destructive	14	14 Prospective	Consecutive	HGG	36.2		7 40 minutes	Histopathology	0.86	1.00
MRI	Colubrier 2017	S-AIA	in vivo	Non-Destructive							10 mins	Histopathology	0.60	1.00
MRI	Colubrier 2014	ioMRI	in vivo	Non-Destructive							not reported	Histopathology	0.66	0.60
Optical	Hauser 2016	S-AIA	in vivo	Non-Destructive		145 Prospective	Selected	HGG			8 40 mins	Histopathology	0.81	0.43
Optical	Mausmeyer 2007	Autofluorescence and Diffuse Reflectance	in vivo	Non-Destructive	83	250 Prospective	Consecutive	Gliomas	NA		9 minutes/ hours	Histopathology	0.96	0.97
Optical	Gong 2008	Elastic light scattering spectroscopy	ex vivo	Non-Destructive	29	36 Prospective	Selected	HGG	NA		8 laboratory analysis	Histopathology	0.8	0.93
Optical	Reigner 2012	Raman Microscopy	ex vivo	Non-Destructive	1	39 Prospective	Selected	Mets	NA		9 hours	Histopathology	0.99	0.97
Optical	Jennings 2015	Raman Probe	in vivo	Non-Destructive	17	361 Prospective	Selected	Brain tumour	53		7 Seconds	Histopathology	0.93	0.91
Optical	Zhang 2017a	Fluorescence	in vivo	Non-Destructive	38	89 Prospective	Consecutive	HGG			7 NA	Histopathology	0.94	0.89
Optical	Ji 2015	SRS Raman Microscopy	ex vivo	Non-Destructive	22	75 Prospective	Consecutive	Gliomas			8	Histopathology	0.975	0.985
Optical	Gessler 2015	S-AIA	in vivo	Non-Destructive	32	32 Prospective	Consecutive	HGG				Histopathology	0.7	1
Optical	Du Le 2017	Diffuse Reflectance and Fluorescence	ex vivo	Non-Destructive	7	22 Prospective	Selected	Gliomas			10	Histopathology	1	0.9
Optical	Iluthe 2011	Fluorescence lifetime spectroscopy	in vivo	Non-Destructive	42	71 Prospective	Selected	HGG			7 1 Prospective	Histopathology	1	0.98
Optical	Iluthe 2011	Fluorescence lifetime spectroscopy	in vivo	Non-Destructive	42	71 Prospective	Selected	HGG				Histopathology	0.47	0.95
Optical	Key-Ding 2014	Fluorescence sodium	in vivo	Non-Destructive	6	26 Prospective	Selected	HGG	63.5		7 37 mins	Histopathology	0.79	1
Optical	Wang 2017	S-AIA & spectroscopic probe	in vivo	Non-Destructive	16	16 Prospective	Selected	HGG	82		7 7 Prospective	Histopathology	0.79	0.33
Optical	Herli 2008	S-AIA	in vivo	Non-Destructive	71	71 Prospective	Selected	HGG			9	Histopathology	0.98	1
Optical	Valdes 2011 80	S-AIA & spectroscopic probe	in vivo	Non-Destructive	10	88 Prospective	Selected	Gliomas			10	Histopathology	0.94	0.94
Optical	Valdes 2011 78	S-AIA & spectroscopic probe	in vivo	Non-Destructive	14	70 Prospective	Selected	Brain tumours				Histopathology	0.84	0.92
Optical	Wulsham 2013	S-AIA	in vivo	Non-Destructive	59	215 Prospective	Selected	HGG	41		8	Histopathology	0.89	0.88
Optical	Zimmermann 2001	inTRPC-mediated POD	in vivo	Non-Destructive	22	138 Prospective	Selected	HGG			9	Histopathology	0.979	0.937
Optical	Pancani 2012	S-AIA	in vivo	Non-Destructive	23	92 Prospective	Consecutive	HGG	range (43-79)		11 N/A	Histopathology	0.91	0.89
Optical	Bruckner 2016	Confocal Laser Endoscopy	ex vivo	Non-Destructive	100	100 Prospective	Consecutive	HGG	N/A		7 Minutes	Histopathology	0.81	0.85
Optical	Bruckner 2016	Confocal Laser Endoscopy	ex vivo	Non-Destructive				HGG				Histopathology	0.90	0.93
Optical	Seibotha 2009	Infrared Spectroscopy	ex vivo	Non-Destructive	6	54 Retrospective	Selected	Gliomas	58.17		8 15 min	Histopathology	0.93	0.88
Optical	Eustachian 2011	Confocal Microscopy	in vivo	Non-Destructive	20	45 Prospective	Selected	Brain Tumour N/A			7 30 minutes	Histopathology	0.93	1.00
Optical	Colubrier 2014	S-AIA	in vivo	Non-Destructive	34	114 Prospective	Selected	HGG	59		7 not reported	Histopathology	0.91	0.80
Optical	Stummer 1998	S-AIA	in vivo	Non-Destructive	59	89 Prospective	Not reported	HGG	58.7		7 25-47 minutes	Histopathology	0.85	1.00
Optical	Lau 2016	S-AIA	in vivo	Non-Destructive	21	211 Prospective	Selected	HGG	53.3		7 N/A	Histopathology	0.82	0.64
Optical	Martinsen 2018	Confocal Laser Endomicroscopy-Fluorescence	Both	Non-Destructive	74	141 Prospective	Consecutive	Gliomas	47.5 years		8 15.7 minutes	Histopathology	0.91	0.94
Optical	Toms 2005	Reflectance fluorescence spectroscopy	in vivo	Non-Destructive	24	230 Prospective	Selected	Gliomas	N/A		8 30 seconds	Histopathology	0.80	0.89
Optical	Zeh 2017	HIS second window	Both	Non-Destructive	10	13 Prospective	Selected	HGG	61.6		9 22.3 hours	Histopathology	0.86	0.25
Optical	Zeh 2017	White light	in vivo	Non-Destructive								Histopathology	0.77	1.00
PET imaging	Floeth 2011	F-FET PET	in vivo	Non-Destructive	30	38 Prospective	Selected	HGG			8	Histopathology	0.88	0.84
PET imaging	Braun 2002	Methionine PET	in vivo	Non-Destructive	32	34 Prospective	Selected	HGG			8	Histopathology	0.87	0.75
PET imaging	Kinoshita 2002	C-methionine PET, F-FDG PET	ex vivo	Non-Destructive	20	54 Prospective	Not reported	Gliomas			10 unclear	Histopathology	0.935	0.875
Ultrasound	Colubrier 2017	Intraoperative Ultrasound	in vivo	Non-Destructive	33	99 Prospective	Selected	HGG	60			Histopathology	0.9	1
Ultrasound	Colubrier 2015a	Intraoperative Linear Array Ultrasound	in vivo	Non-Destructive	13	30 Prospective	Selected	HGG			8	Histopathology	0.79	0.67
Ultrasound	Colubrier 2015a	Intraoperative Ultrasound	in vivo	Non-Destructive	13	30 Prospective	Selected	HGG				Histopathology	0.21	1
Ultrasound	Colubrier 2015b	Intraoperative Linear Array Ultrasound	in vivo	Non-Destructive	20	68 Prospective	Selected	HGG				Histopathology	0.76	0.58
Ultrasound	Rygh 2008	3-D iOUS	in vivo	Non-Destructive	19	186 Prospective	Not reported	HGG	59		9	Histopathology	0.69	0.75
Ultrasound	Ungless 2005	3-D iOUS	in vivo	Non-Destructive	27	139 Prospective	Not reported	HGG	59		9	Histopathology	0.72	0.36
Ultrasound	Yu 2015	Contrast enhanced iOUS & TEM	in vivo	Non-Destructive	100	902 Prospective	Consecutive	Gliomas	39.6		9	Histopathology	0.621	0.928

Table 4. Diagnostic accuracy of intraoperative brain tumour tissue detection.

3.1.7. Conclusions

A wide range of technologies have been tested for brain tumour detection during surgery. Table 5 summarises the pooled diagnostic accuracies together with key features determining the use of each technology.

Intraoperative Technology	Real-time?	Objective ?	Turnaround Time	Tumour Type	Information	Average Sensitivity	Average Specificity
Fluorescein	+	-	Seconds	HGG only	Structure	86.5	94.5
5-ALA	+	-		HGG only	Molecular	86	85
Confocal Microscopy/ Endoscopy	+	-	10-20 mins	All	Structure	88.8	93
Elastic scattering reflectance		-	Unknown	All	Structure	90	91.5
ICG second window	+		Seconds	All	Structure	82	25
Raman	+	+	Seconds	All	Biochemical/ Molecular	96.5	95.5
Flow Cytometry	-	+		All	Molecular	88	88
MRI	-	-	30-60 mins	All	Structure	71	81
US	+	-	Minutes	All	Structure	67.8	76.8
PET scanning	-	+	N/A	Selected	Molecular	89	82.3
Frozen section/ Cytology	-	-	30 mins	All	Structure	89	98
Mass spectrometry: DESI	-	+	30 mins	All	Molecular	96.1	98.6

Table 5. Summary of technologies used for intraoperative brain tumour detection.

Frozen section

Frozen section has long been the standard for intraoperative diagnosis in brain tumour surgery. By its very nature frozen sectioning requires destruction of tissue and is an *ex vivo* method. The specimen requires transporting to the laboratory, preparation and reporting by a pathologist. This is time and resource consuming and cannot be readily repeated several times during surgery. The analysis does not contain any molecular information and is simply a report on structural features of the tissue sample. This method is subject to sampling error and take 30-40 minutes in a well staffed department, and is not available out of hours for emergency work or for difficult cases which overrun.

Cytology

Touch, imprint and smear cytology has been long used for intraoperative diagnostics such that large retrospective studies exist to assess its diagnostic accuracy. The main advantages of cytology are that very small specimens are required and they are faster to prepare compared to frozen section. Cytology requires the presence of a trained pathologist experienced in this technique and is subject to sampling error much like frozen section, does not give molecular information and it still requires a specimen to be labelled and transported to the laboratory for analysis which will be limited outside of normal working hours.

Rapid immunohistochemistry

Rapid immunohistochemistry was investigated by one study (Tanino et al., 2015). Alternating electric fields were used to speed up the antibody-antigen reaction but this still requires extensive tissue preparation, laboratory staff and a pathologist to interpret the result making the shortest turnaround time just over 30 minutes. It is still subject to sampling error and repeated analysis is not easily possible due to the resources involved in turning around a result. It does however show that standard molecular histopathological techniques are evolving over time and becoming faster.

Flow cytometry

Flow cytometry has been developed which was investigated in 2 studies (Shioyama et al., 2013), (Koriyama et al., 2018). With a mean turnaround time of 9.5 minutes the system was used intraoperatively. This technique requires sample preparation with reagents and pipetting of the tissue sample and is able to gather molecular information at the cellular level. Further development of this technology may speed it up further but it remains unable to give information at the time of surgery in real time.

MRI

IoMRI was investigated using a relatively small number of patients (5 studies involving 101 patients) reflecting on its limited use in a handful of centres due to cost. The process of performing an intraoperative MRI scan is not simple, adds significant time to surgery and involves the use of extra staff for patient transfer and operation of the MRI scanner itself. This reduces the number of scans which are feasible during surgery. The studies included in this analysis performed tissue biopsy correlated to MRI imaging and the results were disappointing with an average sensitivity of 71% and specificity of 81%. This in part reflects on the lack of consensus as to which MRI modality best measures extent of disease and the fact that the tissue resolution of MRI is poor, with microscopic disease extending beyond T1 contrast and T2 signal. Although an excellent surgical planning tool, MRI in its current form proves to be a poor tumour detection system to guide surgery when compared with histological analysis. False positive findings were also seen with contrast leakage at the surgical tumour margin and with surgical effects on surrounding tissue (Heselmann et al., 2017). With the use of higher field, ioMRI systems and other modalities such as perfusion and diffusion imaging this may change. All studies reported false negatives using ioMRI. It is possible, therefore, that the effect of increased EOR using ioMRI reported in some studies (Eyupoglu et al., 2013) may be due to simply greater emphasis on obtaining EOR in a trial environment or the updating of neuronavigation with respect to normal brain structures rather than the ability of ioMRI technology to detect tumour tissue *per se*.

Ultrasound

After MRI, Ultrasound remains the only other commercially available technology for intraoperative assessment of tumour tissue. A wide range of transducers are available with only one company (Sonowand) currently offering ultrasound combined with neuronavigation. Ultrasound has always been cost effective, and integration with navigation has increased precision. It is however plagued by the problem of artefacts and inter observer variability which has limited its uptake. Indeed in this review its sensitivity and specificity was one of the lowest in the group of technologies assessed, being worse than intraoperative MRI.

PET Imaging

PET scanning has been used as part of trials to help identify the location of highest malignancy in diffuse or heterogenous tumours to help enable a diagnostic procedure. Methionine PET has a short half life of 20 mins and this technique is limited to those facilities with a cyclotron to make this PET tracer. FET-PET scanning using ¹⁸F was also used and has a longer half life of 109 minutes. Although a good example of the use of molecular information to help target tumour tissue the spatial resolution of PET scanning is very poor when compared to MRI scanning and the

millimetre accuracy of neuronavigation. These factors have limited its use to select cases where a diagnostic biopsy alone is required in a diffuse tumour or one that displays a large amount of heterogeneity.

Optical technologies

Optical coherence tomography, confocal microscopy and endoscopy allow visualisation of the underlying tissue on a microscopic scale, obtaining information on the structure at a cellular level. Molecular information is not obtained. The images obtained require expert assessment, much the same way as traditional histopathological images, with the exception that staining and sample preparation is not required. Moving to a cellular level of detail during surgery can however be quite bewildering as surgeons do not operate at this small scale. The technique does however provide a non-destructive label free technique to identify tissue of any type.

Diffuse reflectance and light scattering spectroscopy achieved accurate results in small *ex vivo* studies but are yet to be applied during surgery. This may reflect difficulty in standardising measurements when used in an operative environment rather than a laboratory environment and may limit the use of these technologies *in situ*.

Fluorescein travels through the plasma and reaches the tumour through a disrupted blood brain barrier, hence its use in GBM. Unfortunately, this can also lead to false positive results during surgery as progressive physical disruption of the blood brain barrier may result in leakage of more fluorescein and has limited its more general use (Stummer, 2015). It is not useful for lower grade gliomas.

Raman spectroscopy was used to gather molecular information and proved to be highly accurate. The main advantage of this technique is that it is rapid (real time), label and preparation free and does not destroy tissue. It can be used repeatedly by the surgeon to assess tissue. The disadvantages involve the use of lasers and the need to reduce infrared contamination of the operating site for *in vivo* systems. The technology is not yet commercially available for tumour surgery.

Although 5ALA has been a welcome addition to the surgical adjuncts available to detect tumour tissue it is by no means a total solution. Use of 5ALA may not allow detection of tumour tissue if it is hidden by healthy tissue, blood, poor exposure through the resection cavity or the retractor. More than just the tumour may fluoresce - surrounding reactive gliosis for example - resulting in false positives. Heterogeneity within GBM may result in false negatives where some areas of tumour fluoresce more than others. There is also of course the fact that fluorescence from low grade gliomas is not detectable even under microscope blue light illumination. Surgeons also cite the difficulty of acclimatising when switching from white light to blue light frequently as is required to visualise structures when operating (going from bright illumination to dark). Add to this the cost implications of administering 5ALA to the patient, the need to time this for surgery limiting operating list flexibility and the risk of side effects from the medication such as skin photosensitisation and headache and all of a sudden it is not such an easy adjunct to use as and when required within an overburdened healthcare system.

Mass spectrometry

This is a destructive tissue technique which analyses tissue at the molecular level to develop a detailed profile of tumour tissue. When used as DESI the technique can only process *ex vivo* tissue and is not rapid enough to be real-time and immediately inform surgery. It does have the potential to have high diagnostic accuracy.

Combination of technologies

Inconsistencies between ioMRI and 5-ALA fluorescence and ultrasound have been observed. Although with all of these technologies false negatives are seen there are relatively few false positives meaning that combining modalities may improve the identification of tumour tissue. Due to the relative strengths and weaknesses of each single modality above a combined modality approach to improve diagnostic accuracy and mitigate the drawbacks of using one modality alone should be pursued. For example, ioMRI may assist in identifying the bulk of residual contrast enhancing disease, after which point-analysis techniques such as Raman may be used to fine tune surgical resection.

It is clear from reviewing the literature that the ideal tissue detection technology remains to be discovered. This technology should be rapid, allowing real-time data processing, be based on molecular information in order to be accurate and specific, and should be easy to use not requiring any tissue preparation, labelling or laboratory processing.

4. Raman Spectroscopy

4.1. Overview

This section deals with the development and application of Raman spectroscopy for the first time as an *in vivo* technique for brain surgery. I overcame a number of challenges including system development, calibration and setup in the operating room. I developed a specific brain tissue data analysis methodology which will be discussed. The results of the first *in vivo* use of the system will be presented using multivariate statistics. My aim is to show how a laboratory technique can be modified for use in an environment such as an operating theatre and yet retain enough accuracy for important conclusions to be drawn from the tissue being analysed.

4.2. Method Development

4.2.1. Introduction

The proposed use of Raman spectroscopy outside of the controlled environment of the laboratory requires careful consideration if it is to be successful in obtaining useful data to characterise human tissue *in vivo*. This chapter explains how I standardised the process of *in vivo* Raman spectroscopy, how unwanted variation between spectra was identified and removed, leading to the ability to take repeated *in vivo* Raman spectra of acceptable quality. This was a necessary first step prior to successful *in vivo* application of the Raman system.

4.2.2. Method Development

In order for *in vivo* Raman spectroscopy to be carried out successfully, a method development phase was used. The purpose of this phase was to decide how to control important variables which can influence Raman spectra and to develop a system which can be deployed in the operating theatre and be used repeatedly. Although extensive preparation was done before the use of the system in terms of selecting laser wavelength and a sterilisation protocol for the system, many aspects of the final surgical protocol needed to be refined during the first use of the system in the human patient. These aspects could not be investigated using frozen tissue and included selecting the correct laser power and controlling background IR *in situ* and accounting for blood in the spectrum.

4.2.2.1. Laser wavelength

A range of laser wavelengths is available in Raman spectroscopy, ranging from 455nm-1064nm in common bench-top systems. Lower wavelengths have higher energy and are more likely to result in fluorescence from the tissue being analysed dominating the spectrum, resulting in a poor Raman spectrum. This is because fluorescence is more quantum efficient (1-3 in 10 photons resulting in fluorescence) than the Raman effect (1 in every 100 photons being scattered). In order to minimise this effect, a longer wavelength laser with lower energy is less likely to excite molecules into a higher energy state that might lead to fluorescence. It is more likely to be scattered as part of the Raman effect. However, the drawback is that Raman efficiency is lower at longer wavelengths resulting in much smaller peaks. Given the lower energy it is also less likely to cause damage to brain tissue.

785 nm laser wavelength was used in order to avoid the problems of fluorescence, and when matched with a super cooled CCD with a superior photon detection capability, smaller peaks could still be analysed. With each laser specific gratings are used. This is one of the reasons multiple laser systems, which would allow a combination of high intensity and low fluorescence, are much costlier to build.

4.2.2.2. Sterilisation

The probe was first passed through several sterilisation tests to confirm safe reusability in the operating theatre. This involved developing a fibre cleaning protocol and gaining Advanced Sterilisation Products (ASP) certification. This required bioburden (or 'microbial limit') testing at the ASP facility owned by Johnson and Johnson, Irvine, CA 92618, USA. This was followed by functional testing of fibre optic integrity at the British Columbia Cancer Institute, Vancouver, Canada. The probe was deemed compatible with sterilisation using the STERRAD system in use at the Charing Cross sterilisation department and was certified by the sterilisation department for use during surgery.

4.2.2.3. Spectral range

The maximum spectral range possible with this system was used - ranging from 300-1900 wavenumbers. The system can only display 700-1900 wavenumber range but records down to 300 wavenumbers. The complete data set was used for characterisation purposes as DNA vibrational modes can be in this range and were deemed to be important for distinguishing tumour from non tumour samples.

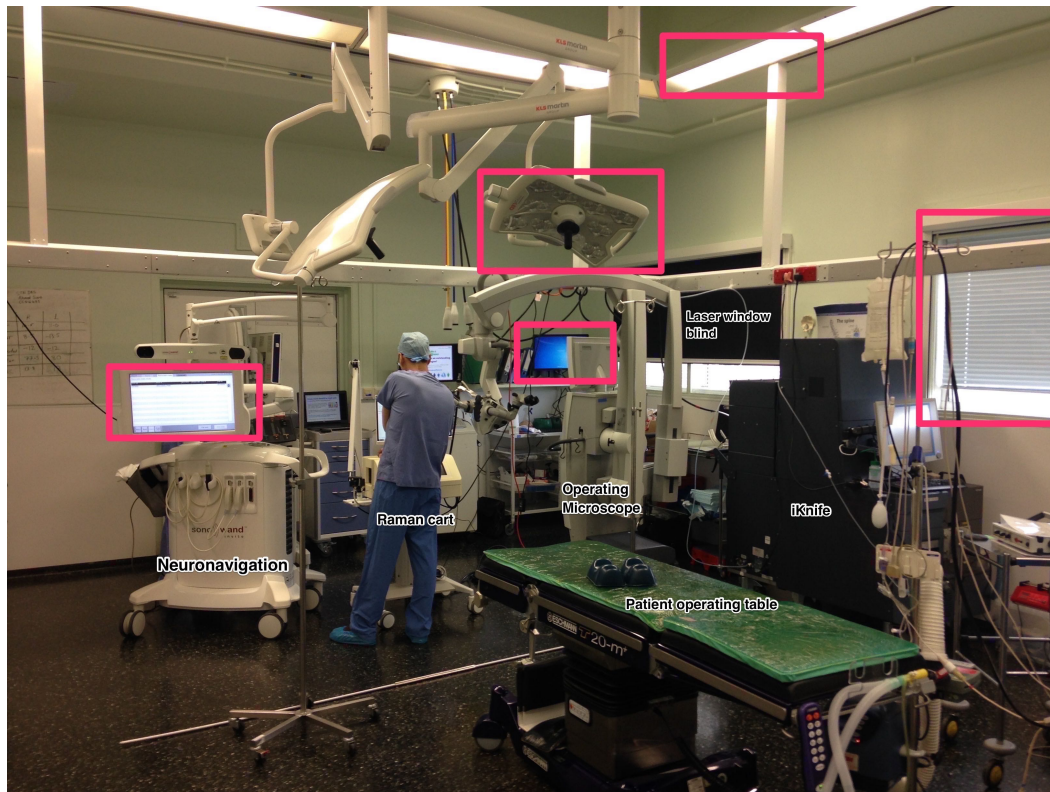
4.2.2.4. Background fluorescence

All significant sources of near infrared (IR) emission that would interfere with the Raman measurements had to be eliminated from the surgical suite, while still keeping the patients safe, not compromising their treatment, and allowing personnel to function as normally as possible (Zhao et al., 2014). This was accomplished by first identifying all sources of significant interference by using a handheld spectrometer (USB2000+VIS-NIR, Ocean Optics Inc., FL, USA). The main sources of interfering light within the surgical suite that I measured were: the main overhead operating light, the surgical microscope with its built in light source, general operating room lights, the neuro-navigation system, and various instrument displays. Interfering light coming from sources outside the surgical suite, and passing through windows and door seams, was found to originate from both natural daylight and artificial light sources. The main sources of IR are shown in Photograph 2 and are summarised as:

1. Sunlight from theatre windows (15th floor: significant sunlight due to lack of obstruction, partially working blinds)
2. Fluorescent theatre ceiling lighting (7 lights)
3. Operating microscope light (halogen light) directed at the operating site
4. Sonowand navigation IR source directed at the patient
5. LCD instrument displays: monitor screens, medical device displays directed to users

Once identified a combination of switching off unnecessary light sources, completely blocking or filtering others and altering the aim of directional light sources was used

to reduce the interference to acceptable levels. For light (natural and artificial) entering the surgical suite through windows and doors, the simplest and most effective approach was to block the light with temporary blackout blinds, opaque door sweeps and draft excluders. Within the operating room itself all unnecessary light sources were switched off and additional floor lamps, with absorption filters to remove interfering near IR emissions, were provided to improve lighting in some locations.



Photograph 2. Theatre 7 setup highlighting sources of infrared.

Sources of infrared light are highlighted using red boxes. These sources include monitor screens, the neuronavigation system, windows and fluorescent lighting. Equipment used during surgery is labelled. The equipment has been positioned ready for the patient to be brought into the operating theatre for surgery.

4.2.2.5. Interference at the operative site

It was critical for the surgical area to be illuminated at all times during the procedure (including while taking Raman measurements of the tissue). Since it was impractical to filter the main overhead operating light, I decided to use the surgical microscope and filter its built in light source to illuminate the tissue during Raman measurements. This allowed for continuous visual monitoring of the Raman probe position. Whereas the main overhead operating light, which had a highly directional emission and was mounted on a pivot with three degrees of freedom, was turned away during Raman measurements. Similarly, the neuro-navigation system which keeps track of the position of surgical instruments in 3D by monitoring the scattering of near IR light was briefly turned away during Raman measurements.

The high intensity of the microscope light source required a high performance filter that could block the interfering near IR emissions, and withstand the elevated temperatures, but still allow visible wavelengths to pass without significant attenuation. A hard coated short pass (SP) interference filter (Edmonds Optics Ltd, York, UK) was chosen for this purpose. The filter had a transmission of $\leq 0.01\%$ (OD4) for all wavelengths from 775 to 1125 nm (note: the wavelength equivalent of the measured Raman range was from 800 to 931 nm) while allowing 90% transmission in the visible from 400 to 740 nm. The filter was placed in a custom designed and fabricated enclosure that could be mounted in front of the microscope light source focusing lens. That lens was found to produce a roughly collimated beam, and thus the illumination light was at approximate normal incidence to the filter which maximized its attenuation properties. Tests showed that there was no significant interference of the Raman spectra after applying this filter.

4.2.2.6. Light Interference from instrument displays

The last significant sources of interfering near IR light were the instrument displays, which various surgical staff had to monitor throughout the procedure. It is well known that many early generation liquid crystal displays (LCD) produce significant near IR light due to their broad spectrum backlights. Whereas light emitting diode (LED) type displays, or new generation LCD displays with LED backlighting, are usually free of IR emission. However, changing critical displays to the LED type is an expensive solution and optical filtering was not practical. I eliminated the interference that the displays caused by turning displays that ancillary workers have to monitor away from the surgical location. For those displays that the surgeons monitor, a procedure was developed into the acquisition of Raman spectra where personnel would briefly rotate the devices (90 degrees away), acquire spectra, and return them to their normal position afterwards.

4.2.2.7. Ambient background residual

The measures described above for reducing the IR light interference on the Raman spectra were largely effective. However, there will always be some residual light interference with open surgeries. There was also the possibility of a transient increase in that residual (e.g. increased reflection from sources aimed away from surgical area due personnel movement). To account for this reality the final method we introduced to account for these likelihoods was to always take a background measurement with the Raman system, but with the laser off. This background measurement was immediately followed by a Raman measurement (with the laser on) of the same site, and thus the difference between these two measurements should be largely free of background effects.

To illustrate the effects of background IR for all the above sources Raman spectra with obvious artefacts from the IR environment I measured in the operating room are demonstrated in the Figure 15. below. This figure shows background fluorescence

reading in blue (laser is off), Raman reading at full power in red, and the difference spectrum in green.

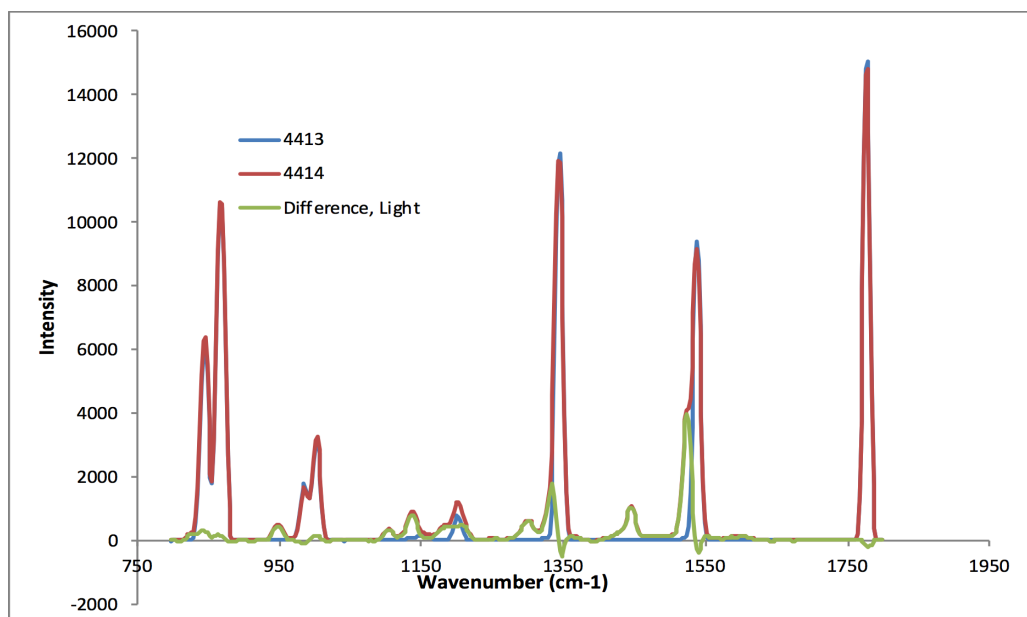


Figure 15. Fluorescence from artificial sources.

The operating microscope light is seen as clear spikes throughout the measured spectrum, orders of magnitude greater than the weaker Raman spectrum. This has the potential to dwarf the Raman spectrum and interfere with Raman readings.

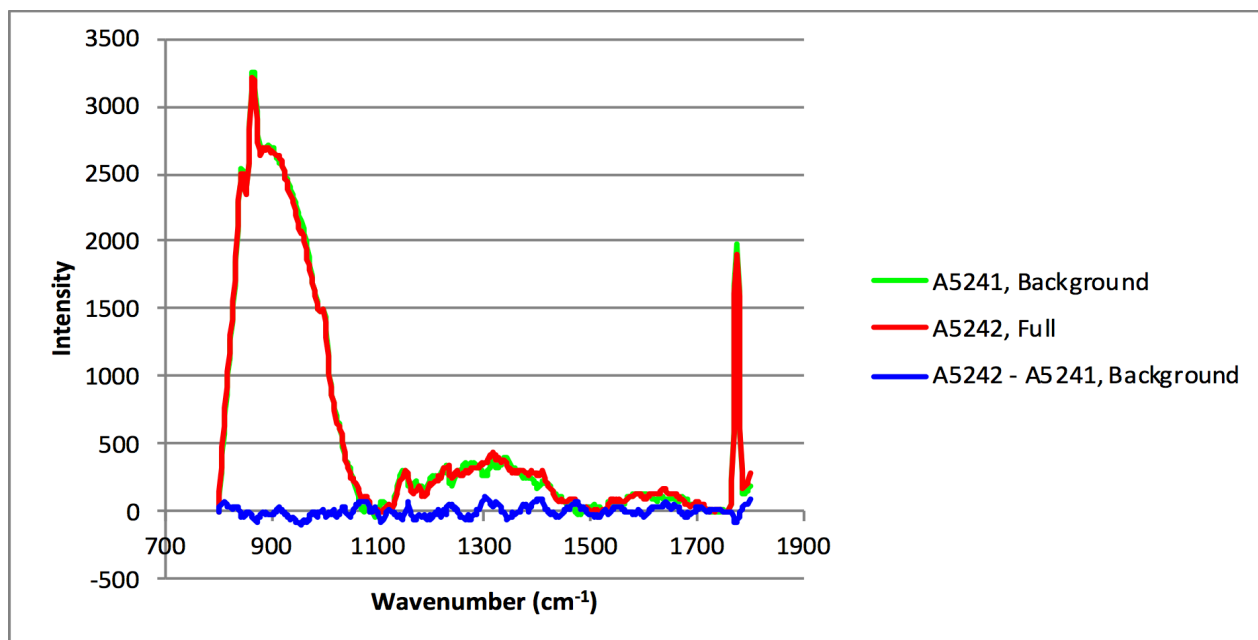


Figure 16. Raman spectrum obtained in the operating theatre on brain tissue with significant fluorescent light background from mixed IR sources.

The green background measurement matches the full Raman measurement (in red) showing that nearly all the signal measured is from the background. In particular the peaks at 850 cm⁻¹ and 1800 cm⁻¹ match the microscope fluorescence spikes above. The poor signal to noise ratio renders this Raman spectrum unusable.

The degree of IR contamination depended on distance from the probe to the IR source which varied according to patient position, microscope and surgeon position, all of which were highly variable. Figure 16 shows the in vivo effect of IR contamination at the operating site. I decided to exclude or control all sources of background IR as identified above in order to ensure the best signal to noise ratio. This involved

1. Sunlight exclusion from the operating theatre - laser blinds placed on all windows
2. Ceiling fluorescent lamps to be switched off during Raman spectroscopy
3. LED lighting to be used (negligible IR emission)
4. Optical filtering of the microscope
5. Directing Sonowand navigation IR source temporarily away from the patient, or replacement with Axiem induction coils (this does not use IR to track surgical tools)
6. LCD screens to be switched off, medical device displays to be turned away from the operating site (anaesthetic displays could not be switched off for safety reasons and were turned toward the anaesthetist)

Once identified a combination of switching off unnecessary light sources, completely blocking or filtering others, and altering the aim of directional light sources was used to reduce the interference to acceptable levels. For light (natural and artificial) entering the surgical suite through windows and doors, the simplest and most effective approach was to block the light with temporary blackout blinds, opaque door sweeps and draft excluders. Within the surgical suite itself all unnecessary light sources were switched off and additional floor lamps with absorption filters to remove interfering near IR emissions, were provided to improve lighting in some locations.

The microscope I tested for IR emission was the Leica M530 OH6. An Edmunds Optics #84-782 was fitted onto the microscope under the sterile drapes and allowed all visible wavelengths to pass but cut off ($\geq OD4$) and blocked any transmission in the range 775 – 1125 nm, dealing with the main IR contamination at the lower end of the Raman spectrum. Careful control of sunlight and fluorescent ceiling lights eliminated the strong spikes due to fluorescence in Figure 17. By pointing the Sonowand IR navigation camera away at the time of Raman acquisition I reduced the background IR contamination to near zero.

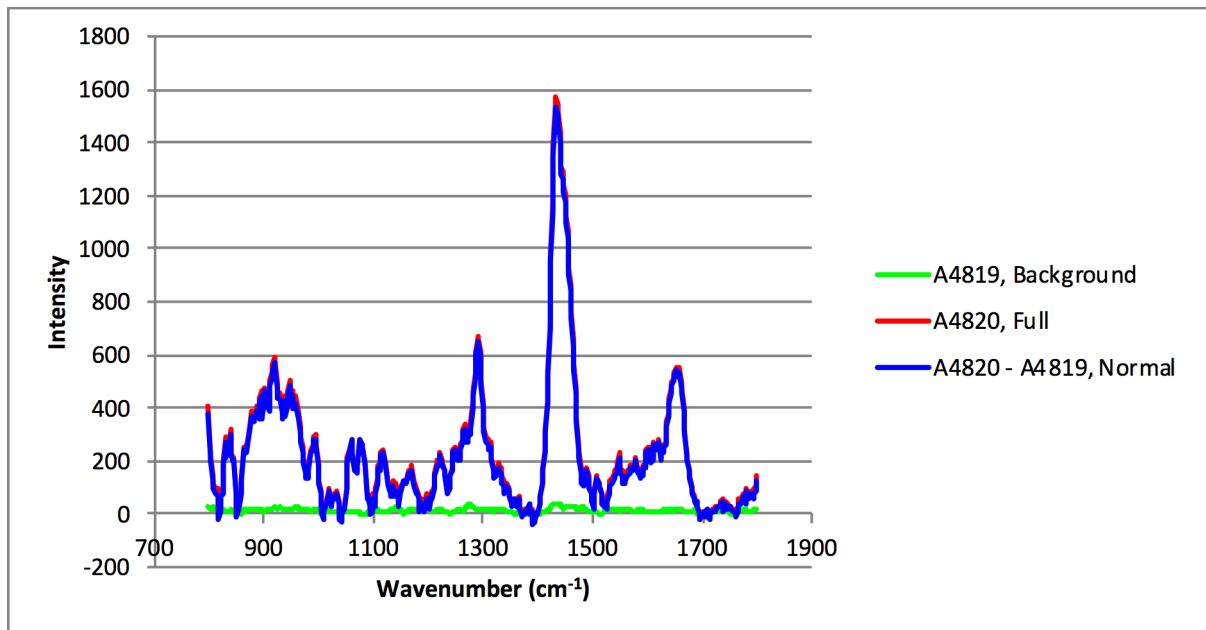


Figure 17. Raman spectrum of the human brain taken during surgery with IR precautions.

Above: Raman spectrum of right frontal lobe of brain, with all IR precautions. Notice the 850cm⁻¹ peak and 1800cm⁻¹ fluorescence peaks are now gone.

4.2.2.8. Laser power and distance from tissue

The system took 2 spectra each time the foot pedal was pressed: first was an ambient light background spectrum with the laser off, and the second was taken at full laser power, (programmable up to a maximum of 150 mW). The maximum excitation laser power was determined by assuring that the power density at the tissue surface is within the ANSI standard for the maximum permissible exposure limit of skin (America, 2014). There is a direct relation between signal intensity and laser power such that the highest power was selected to improve signal quality. Distance from tissue has a similar effect as laser power: the further the probe is from the tissue being analysed, the lower the amount of excitation laser power able to reach the tissue (following the inverse square law). In practice the best signal I obtained was when the probe was a few mm away from the tissue being analysed as shown in Figure 17. This gave a higher intensity Raman signal and also reduced the amount of background IR picked up by the probe as it was closer to the tissue rather than external IR sources. 2mm was therefore defined as the optimum distance from tissue to take spectral readings. In practice this was judged as holding the probe just over brain tissue without touching it (the view through the operating microscope made such positioning possible).

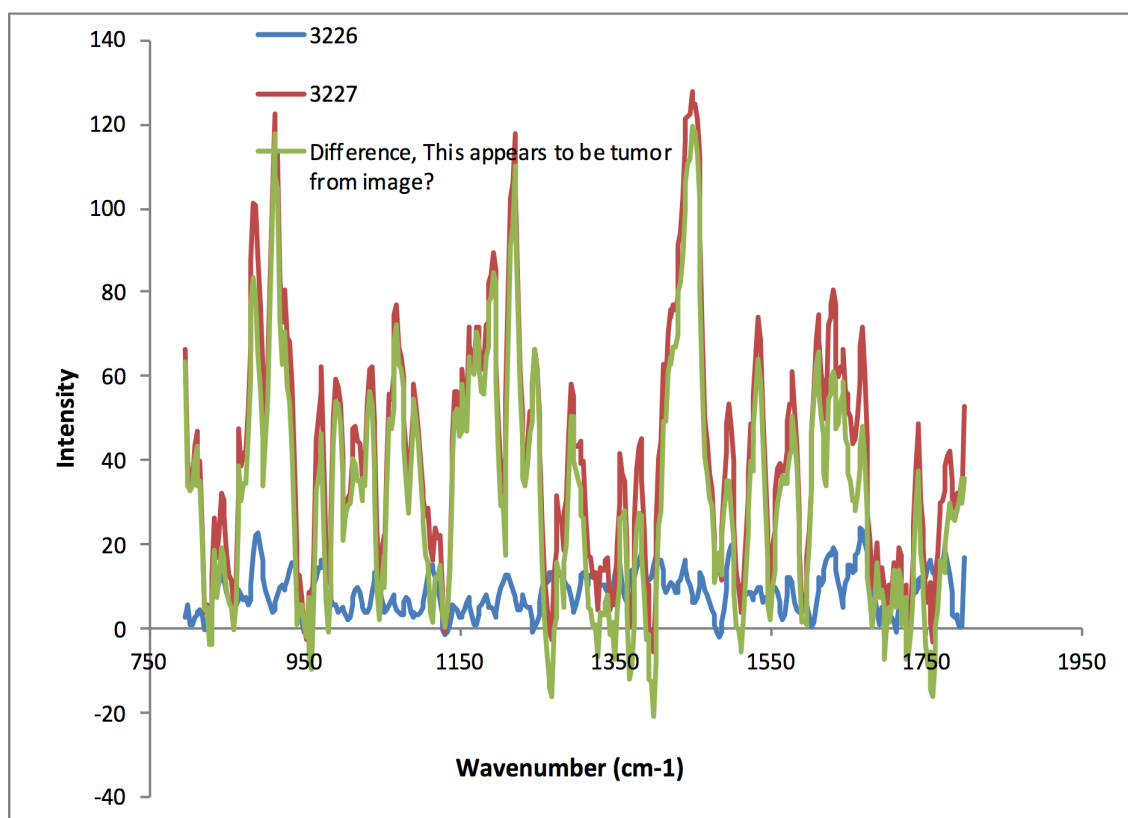
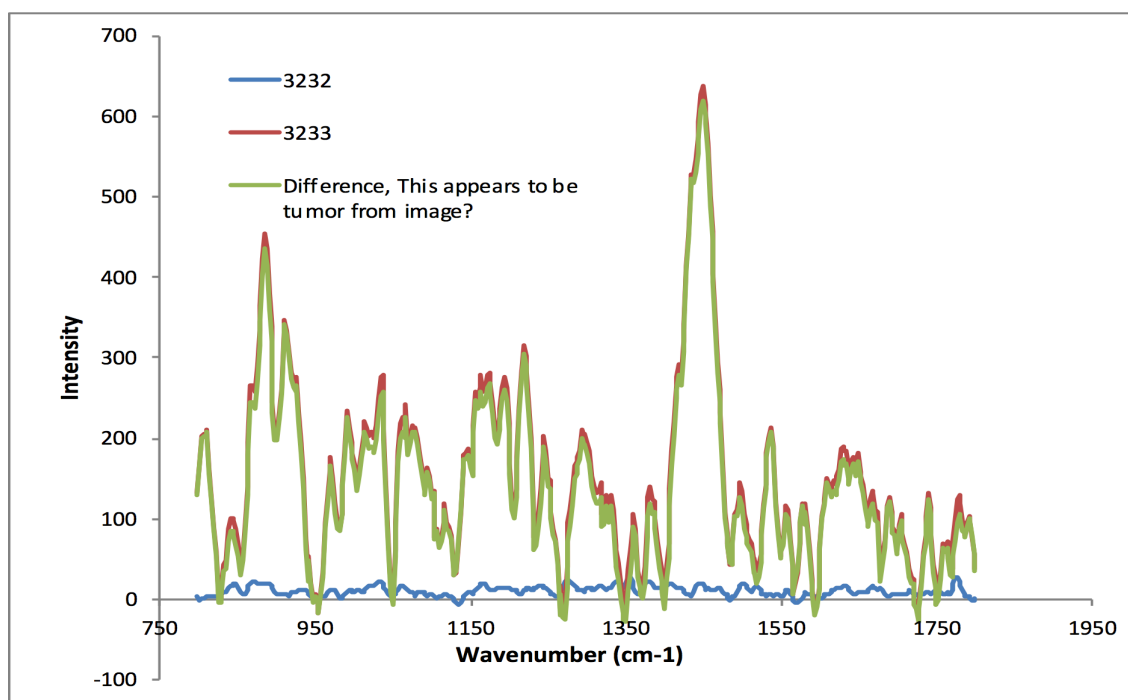


Figure 18. Raman spectra from patient 1.

This shows high intensity and low background IR when spectra are taken 2mm from tissue (top) compared to 1/6 of the intensity and much higher background IR when taken 4mm from tissue (bottom).

4.2.2.9. Blood artefact

I observed the presence of blood clearly in our first patient. This is significant as the human brain is a vascular organ, and many tumours exhibit hypervascularity and can bleed during surgery. This cannot be measured in *ex vivo* tissue and the degree of contribution of the blood signature to brain tumour spectra is unknown. During surgery in our first patient, spectra were obtained in areas of the tumour which were bleeding and in areas in which bleeding had been controlled. Figure 19 shows these spectra and confirms significant peaks are due to the presence of blood.

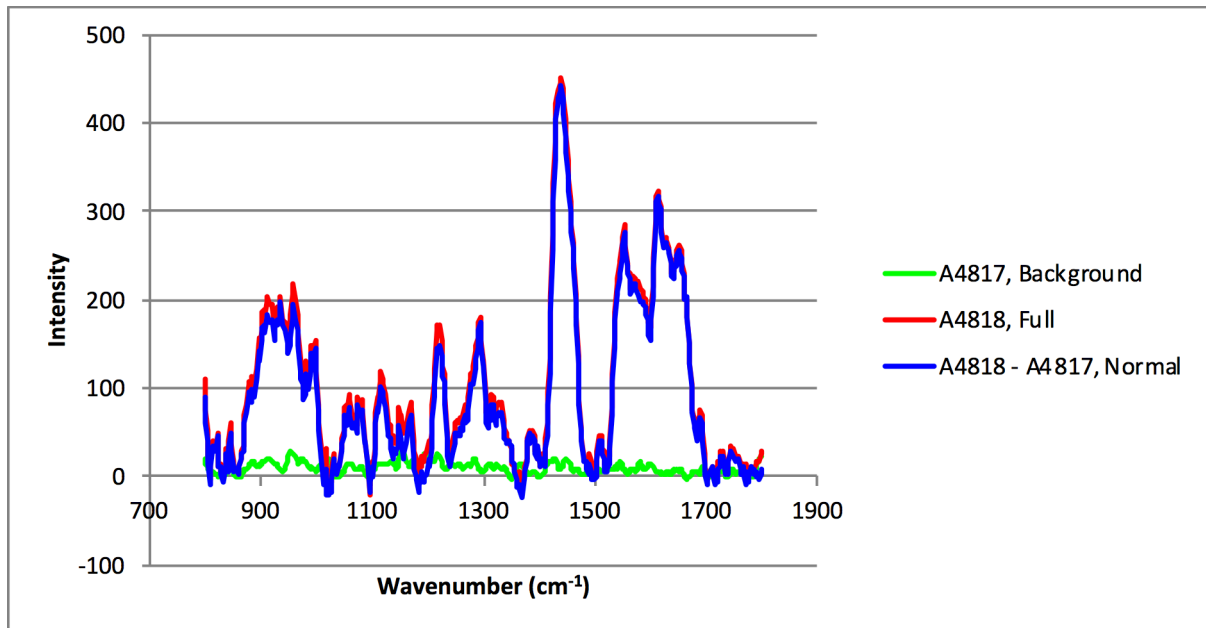


Figure 19. Example of Raman spectrum with blood artefact.

The peaks at 1552 and 1614 cm⁻¹ match known blood peaks, and the presence of this double peak was the most noticeable sign of blood signatures dominating the Raman spectrum, although other peaks are also attributable to blood. Spectra numbers A4817 and A4818 were used for this figure.

These peaks were noted and any spectra showing blood peaks were re-taken after control of bleeding to minimise this effect as shown in Figure 20 which shows a Raman spectrum obtained after haemostasis in the same tissue position as the spectrum shown in Figure 19.

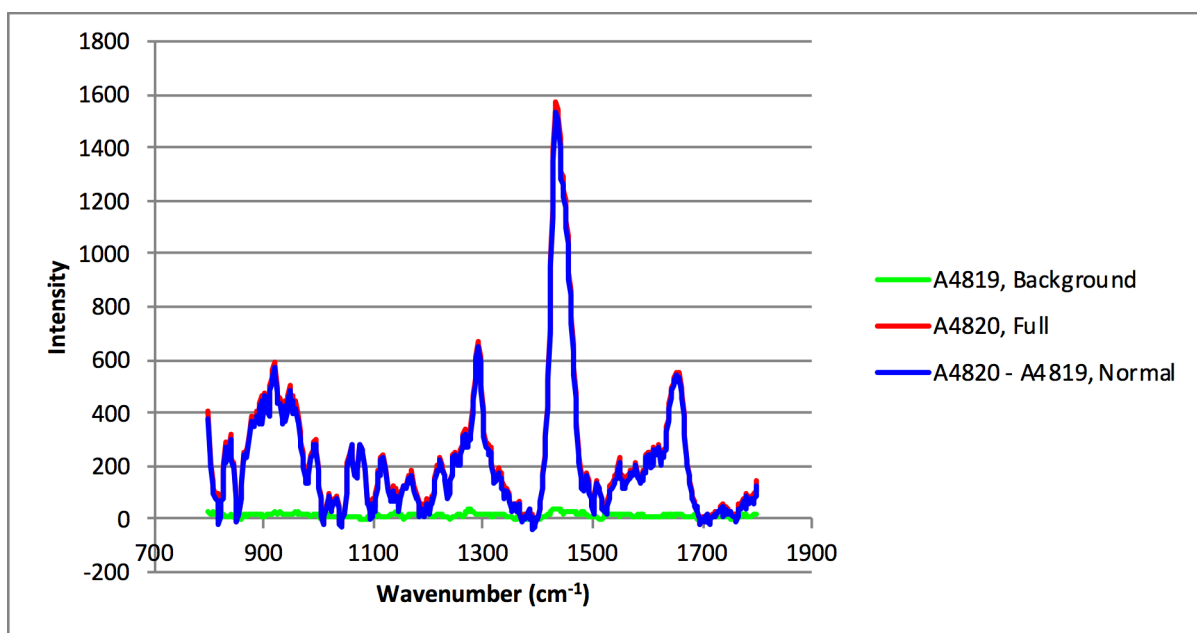


Figure 20. Raman spectrum taken after haemostasis.

Note the reduction in blood artefact signal. This leaves a cleaner Raman spectrum for analysis with many more peaks revealed. Spectra numbers A4819 and A4820 were used for this figure.

The method development work I undertook above resulted in a repeatable method of taking Raman spectra, with standardised attempts to control factors which influence the spectra recorded. This directly led to me guidelines for good quality spectra.

4.2.2.10. Guidelines for good quality spectra

The following 3-point checklist criteria were established as guidelines for good quality spectra:

1. Absence of Fluorescence signal as judged by background readings taken with laser off
2. Absence of Blood signal: peaks at 1550 and 1610 used to identify presence of blood according to in vivo data collected during surgery
3. Signal intensity high: aim for highest peaks to be over 1000 arbitrary units if possible.

4.2.3. Conclusion

A number of variables influencing *in vivo* Raman spectra in the operating room were identified and successfully controlled. IR sources from medical equipment and natural and artificial light sources were the most important factors to take into account. Use of an operating microscope optical filter was crucial in order to enable illumination of the operation site when taking Raman spectra with the probe, both in terms of safety and recording the exact position where spectra have been recorded.

Now that it has been demonstrated that accurate Raman spectra taken in the operating theatre are possible, the next step is to ensure readings can be repeated through standardised calibration and use of the system. Further specific information relating to system calibration and setup follow in the next few chapters.

4.3. System Calibration

4.3.1. Introduction

Since all Raman systems will measure the same emission input differently, each system must be calibrated so that sample emissions can be compared across different systems. The challenge which has limited the use of Raman *in vivo* systems has been standardisation of use. Although the patient spectra present here were not acquired in a multicentre clinical trial, accurate calibration allows for the spectra to be compared to those in the literature. Furthermore, accurate calibration can have an impact on the interpretation of the results.

4.3.2. Calibration

4.3.2.1. Wavelength calibration

Wavelength calibration was performed by using the system to measure the output of two standard lamps Mercury/Argon (Hg/Ar) and Krypton (Kr), both supplied by Ocean Optics Inc., Dunedin, FL, USA. The wavelength position of the peaks in these standards were provided either by the lamp manufacturer, or by looking up the National Institute of Standards and Technology (NIST) database of strong emission lines. The mathematical relationship between the pixel column numbers of the CCD and the known peak wavelengths of the standard lamps was determined using coefficients derived from a 2nd order polynomial fitting to the data. This relationship was then used to wavelength calibrate all columns of the CCD (see Figure 21). This wavelength calibration can then be converted into wavenumbers relative to the laser wavelength (or the equivalent term: Raman shift).

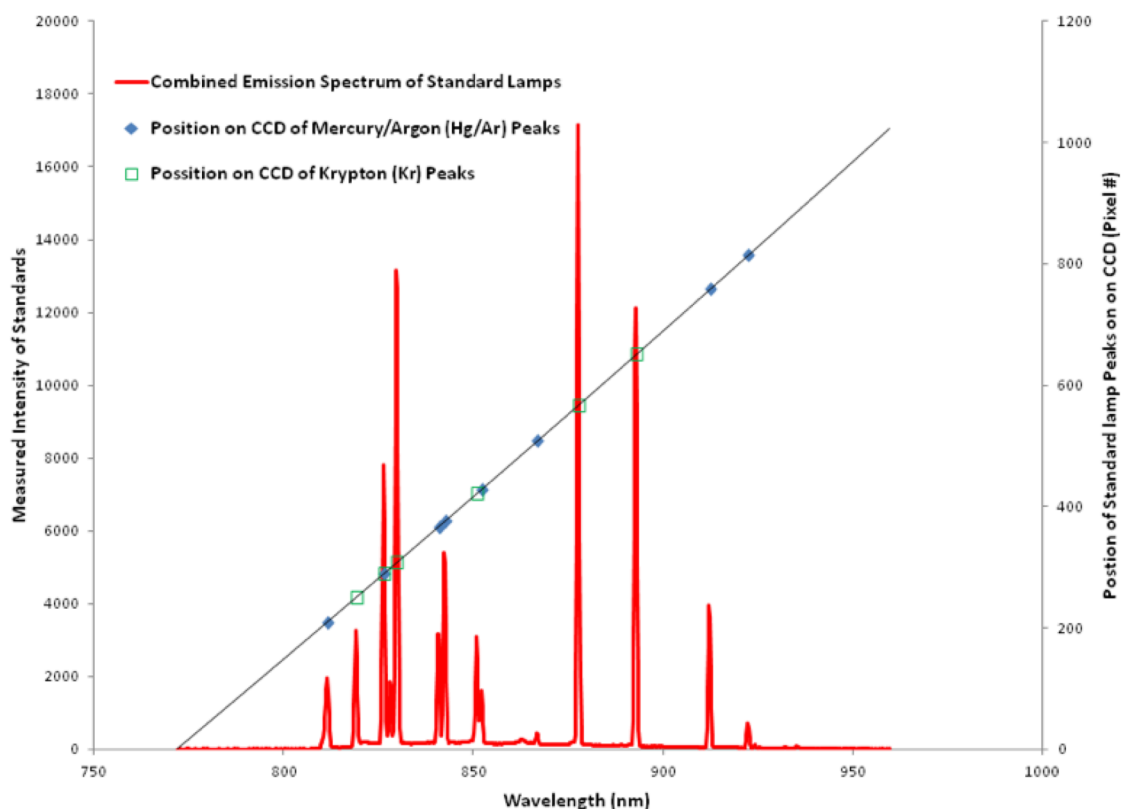


Figure 21. Wavelength calibration of the Raman system.

The red curve shows the combined emission lines from Hg/Ar and Kr standard lamps. The symbols show the peak positions fitted by a 2nd order polynomial, wavelength has been converted into wavenumbers relative to the laser line.

4.3.2.2. Intensity calibration

The intensity response of the system was performed by using the system to measure the output of two broad band emitting standards. The first standard was a calibrated halogen light source (Gamma Scientific, San Diego, CA, USA) coupled to an integrating sphere and the second was a NIST fluorescence standard (SRM-2241). Each standard was measured 10 times and the average calculated. Similarly, the ambient background of the darkroom was measured 10 times and the average calculated. The difference is then plotted and compared with the known intensity output profile of the standards. Both standards gave comparable results. Figure 22 illustrates the response of the system to the SRM-2241 standard. Intensity calibration correction coefficients were then calculated for every data point in a spectrum using the equation:

$$C_{\lambda} = \frac{\text{Known Standard Emission Intensity}_{\lambda}}{\text{Measured Standard Emission Intensity}_{\lambda}}$$

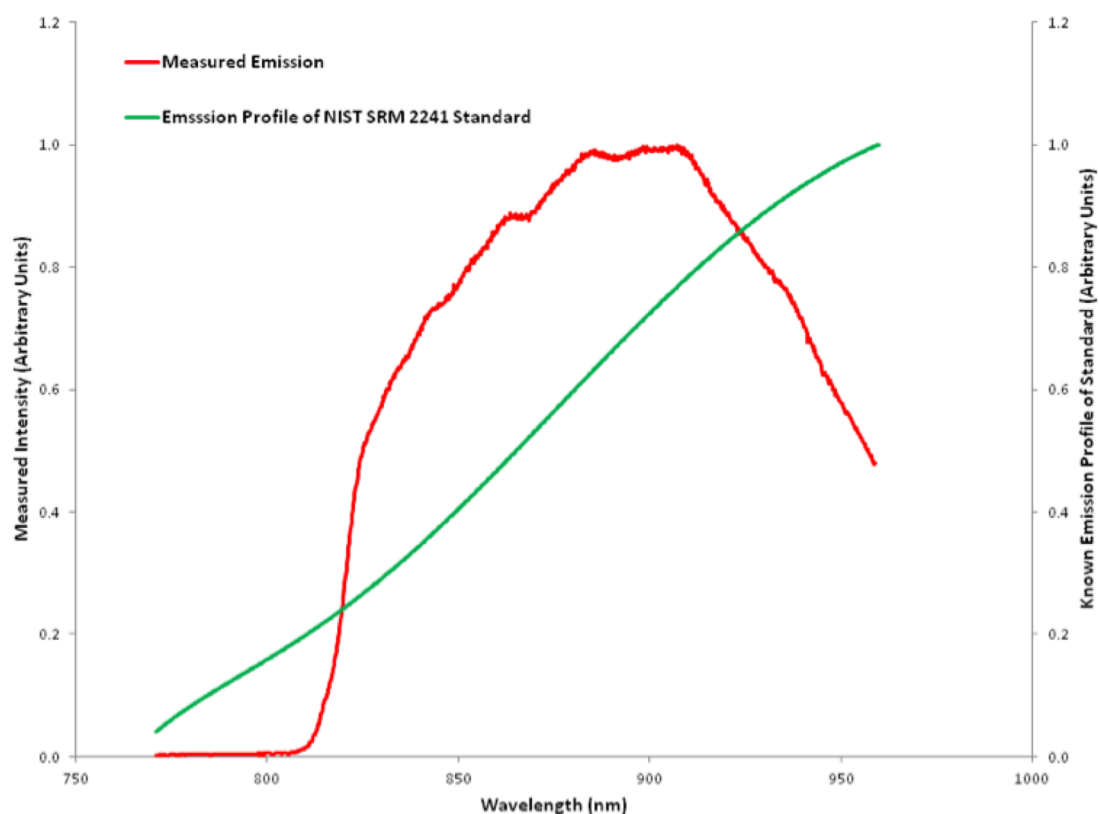


Figure 22. Intensity calibration of the Raman system.

The red curve shows the measured intensity of a NIST, SRM-2241 fluorescence standard. The sharp drop off below 650 cm^{-1} is due to the incorporation of the long wavelength pass filter in the light path to block elastically scattered laser light. The green curve shows the actual emission from the standard (data provided by standard manufacturer).

4.3.2.3. Calibration checks

In addition to the initial calibration setup of the system, the system calibration was rechecked when new patient data was acquired. The rationale for this was to correct for any changes in system components over time. Most changes are expected to be relatively small, however the detachable probe must undergo a rigorous disinfection reprocess, which may result in two potential causes for a significant change in the response of the system. One is that the quick release optical connection may not align exactly every time it is removed and reattached to the system. The second is that the chemical treatment may over time change the optical characteristics of the probe. Although a probe of the same design was tested using 40 reprocessing cycles prior to clinical trials to ensure that it could a) meet the disinfection requirements, and b) it did not suffer from significant mechanical or optical degradation, small changes in optical properties cannot be ruled out entirely.

Thus the calibration of the system was checked by measuring low pressure Hg/Ar and Kr lamp emissions and the SRM-2241 fluorescence standard each time new

patient data was acquired. These data were saved along with the patient data. Later the low pressure Hg/Ar and Kr lamp emission measurements were compared with the original wavelength calibration. No appreciable differences were detected. The SRM-2241 fluorescence standard measurements did indicate there were slight differences in the intensity calibration coefficients, although these did not show any significant trends, and were thus most likely caused by changes that occur in the optical coupling every time the probe is disconnected and reconnected to the system.

4.3.2.4. Interpretation of non-intensity calibrated spectra

It is sometimes argued that if all spectra are acquired by the same system, then analyses of those spectra to look for features that may correlated with independent sample properties should not be affected by the system response. Since both wavelength and intensity calibration involves multiplying the un-calibrated spectra by a series of constants, it is easy to see the basis for this argument. However, apart from the fact that it makes it difficult to compare the spectra with those in the literature, there are also other significant problems. Systems can change with time and thus without calibration checks there is uncertainty whether the measured response is due solely to the sample properties or not.

Secondly, in the presence of significant sample fluorescence, the iterative polynomial fitting only removes the slowly varying component of the measured emission, leaving a residual that scales with the initial fluorescence intensity. This effect is easily demonstrated. In Figure 23 the system response, as derived in Figure 22, is used to generate the measured emission from a series of hypothetical sample emissions. These emissions increase in intensity but the intensity of each member of the series is independent of wavenumber. If these measured emissions are fitted with iterative polynomials (dashed curves) to remove the slowly varying components, the residuals scale with the initial intensity (Figure 24). Although this effect has been demonstrated here for an emission that is independent of wavenumber, it is quite easy to see that this extends to virtually all fluorescence emission profiles. Conversely most fluorescence profiles are slowly varying with wavenumber and are removed successfully by using iterative polynomials on the calibrated spectra. By not calibrating the emission spectra one is effectively introducing peaks into an otherwise slowly varying fluorescence profile, and thus the residual after iterative polynomial fitting is a mixture of both fluorescence and Raman. Whether such spectral mixtures correlate with sample characteristics will depend on the correlation of the fluorescence and Raman components separately and weight each of the emissions have in the analyses method used.

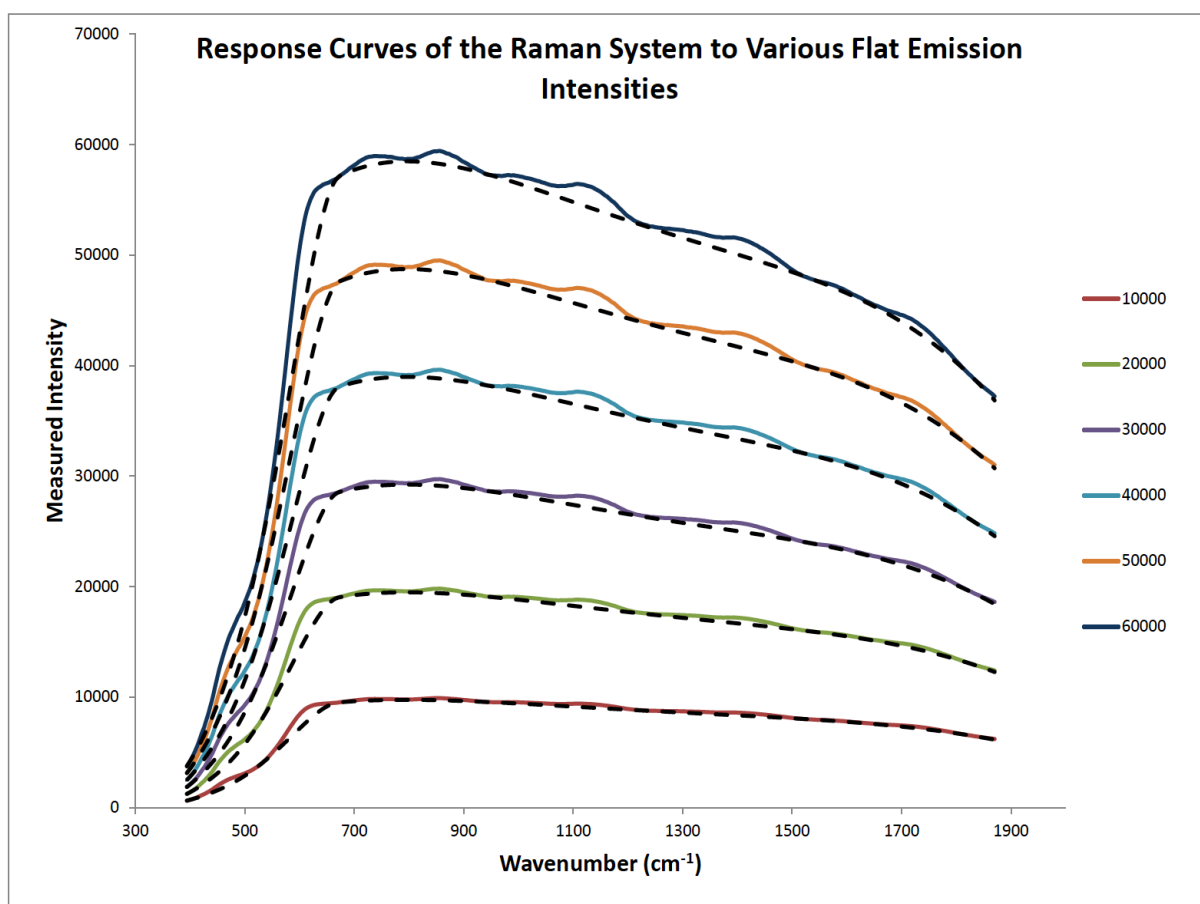


Figure 23. Hypothetical flat emission profiles at 1, 2, 3, 4, 5, and 6 x 10⁴ counts measured by the Raman system.

The dashed black lines are the iterative polynomial fits to the measured emissions. These fits were done in two ranges 400-670 and 672-1870 cm⁻¹ because it is not possible to accurately fit a polynomial to the whole range.

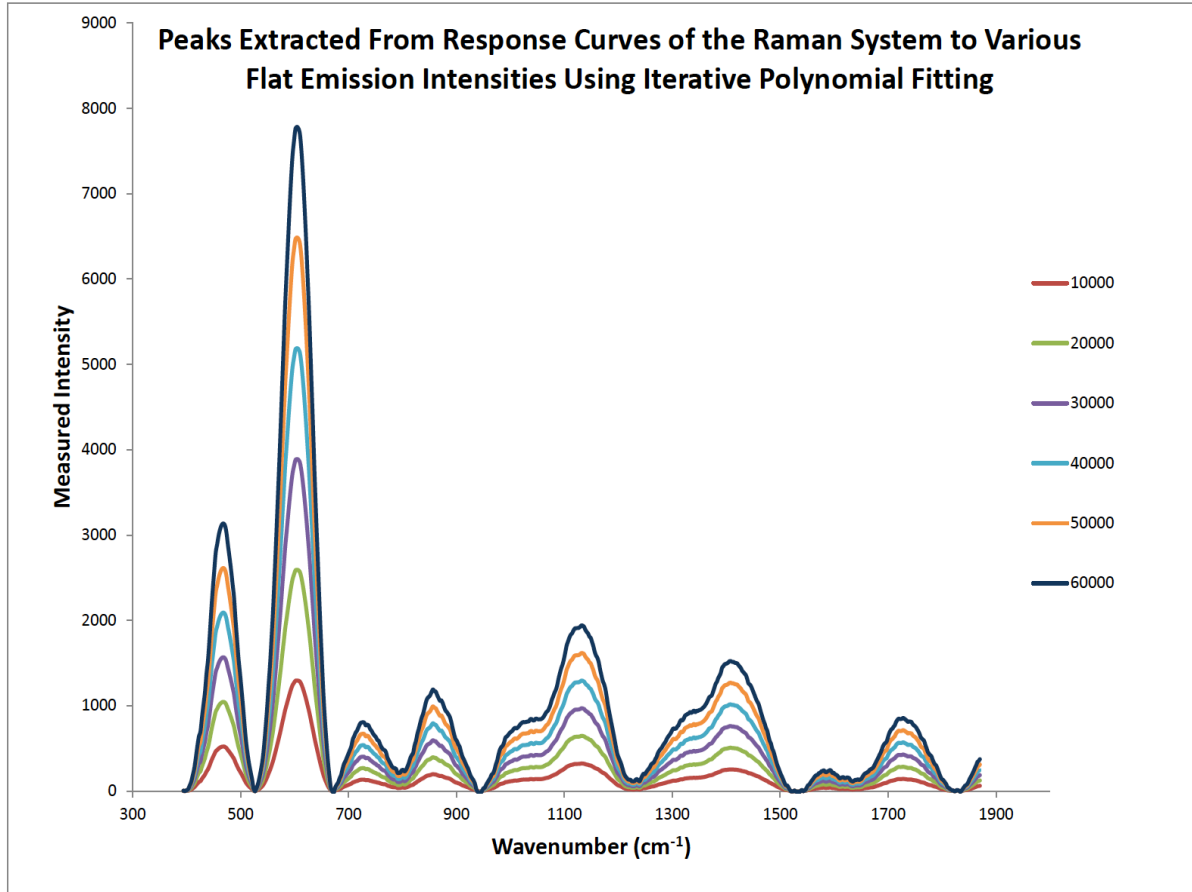


Figure 24. Residuals after subtracting the intensities derived from iterative polynomial fit to the measured emissions of hypothetical flat emission profiles at 1, 2, 3, 4, 5, and 6 x 10⁴ counts.

In Figure 25 the average spectra I obtained are shown for the whole patient data set in the case where the spectra were not calibrated for the response of the system. As expected, the average spectra are very different from the average calibrated spectra since there is significant fluorescence from the sites measured. Also shown are the peaks in the system response which align very well with the dominant peaks from the patient spectra. It is clear that fluorescence is making a major contribution to the spectra.

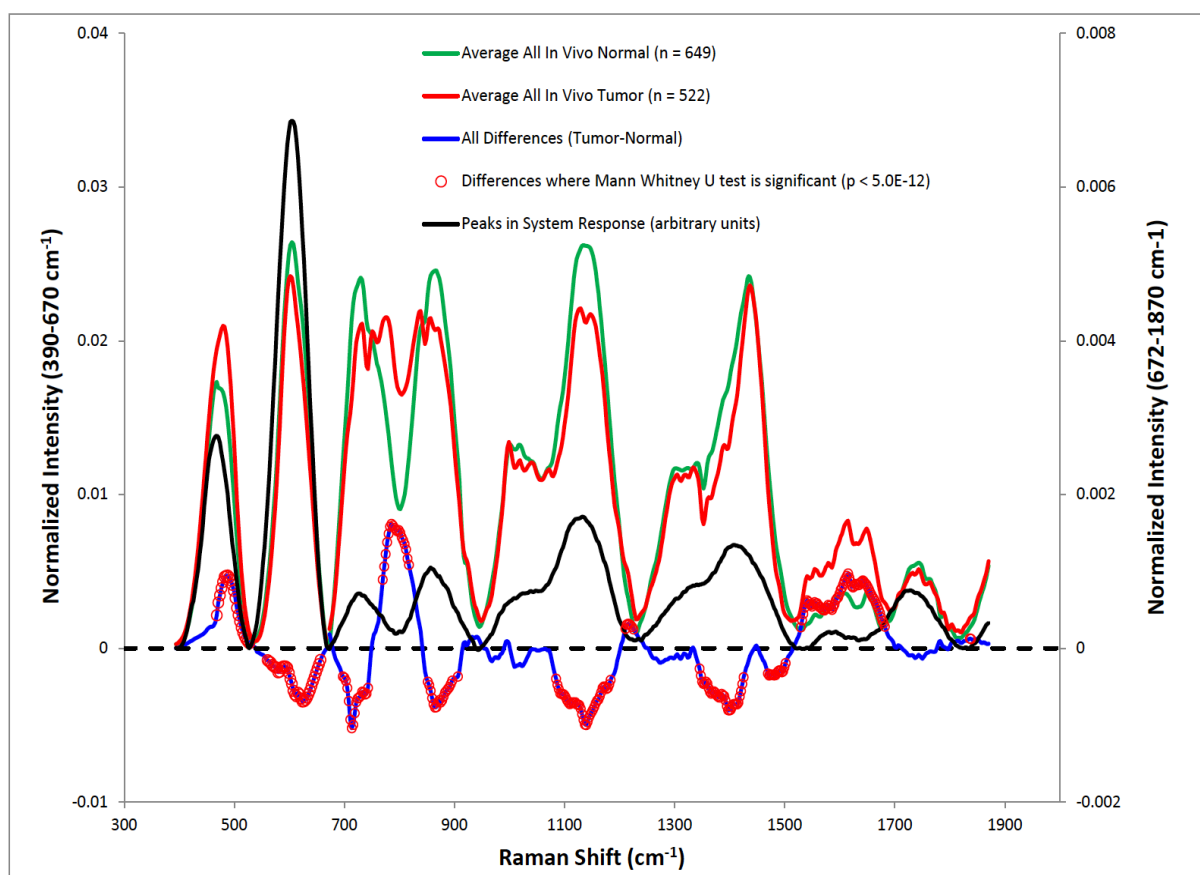


Figure 25. Average tumour, normal and difference spectra using data from all patients.

Raw data was background corrected, smoothed, but not calibrated for the response of the system. Then iterative polynomial fits were done to remove the slowly varying emission component. These fits were done in two ranges 400-670 and 672-1870 cm^{-1} because it is not possible to accurately fit a polynomial to the whole range. The resulting residuals were then average by pathology class. The red circles superimposed on the difference spectrum are positions where there is a statistically significant difference between tumour and Normal averages using a Mann-Whitney U non-parametric test. Also shown are the peaks in the system response (black line).

4.3.3. Summary

This chapter has outlined the steps taken to ensure Raman spectra were appropriately calibrated when the system was used *in vivo* in order to obtain accurate, repeatable measurements enabling data from one patient to be compared with data from another. This was essential before the *in vivo* system could be deployed. The data collected shows that *in vivo* system calibration is possible. Although the precautions taken here were lengthy it is hoped that with future systems calibration will be automated, reducing the time taken to prepare the system for use.

The next chapter outlines the *in vivo* system setup for collecting data during surgery.

4.4. System Setup

4.4.1. Introduction

This chapter details the setup of the system as used in the operating theatre. The components of the system will be discussed including the probe, optical fibres, optical filters, system mounting, system configuration and control.

4.4.2. System overview

A commercial Raman system (Verisante Core) was provided by Verisante Technology Inc. Richmond, B.C., Canada. A schematic of the Raman system is shown in Figure 26. The design was based on similar systems used for *in vivo* studies on human skin, lung, colon, and oral cavity (Huang et al., 2001), (Short et al., 2008), (Short et al., 2016), (Guze et al., 2009). The Raman excitation source was a stabilized 785 nm butterfly diode laser pigtailed to a 105 μm diameter fiber (PD-LD, NJ, USA) and driven by a Verisante designed control module. Tissue emissions were spectrally analyzed with a Verisante, in-house designed, $f/2$ spectrograph incorporating a holographic transmission grating. The intensity of the dispersed emission was quantified with a front illuminated, thermo-electrically cooled (-20°C), charge coupled device (CCD) array with 1024×1024 pixels (Apogee F6, Andor Technology Ltd, Belfast, UK). The size of each CCD pixel was $24 \times 24 \mu\text{m}$, resulting in a total chip size of $24.6 \times 24.6 \text{ mm}$. The system was capable of measuring Raman spectra with shifts of up to 2000 cm^{-1} at a maximum spectral resolution of 7.7 cm^{-1} for a spectrograph slit width of $100 \mu\text{m}$.

4.4.3. Reusable flexible Raman probe

A specially designed flexible bifurcated probe was used for delivering the excitation light to the tissue surface and for collecting the emission from it. The probe was 1.75 m long, consisting of ultra-low hydroxide (OH) optical fibres hermetically sealed in synthetic polymer sheaths and stainless steel terminations. The common end was 2.8 mm in diameter and consisted of a centred 200 μm diameter excitation fibre surrounded by 34, 100 μm diameter collection fibres (not all fibres are shown in Figure 26). All fibres had numerical apertures (NA) of 0.22, and the excitation fibre had a gold coated jacket to prevent cross-talk with the collection fibres. At the proximal end of the probe, the fibres were separated from the bundle to form excitation and emission channels that were coupled to a filter module containing collimating lenses and filters. The coupling was made by way of quick release, sub-miniature A (SMA), optical connectors which allowed the flexible probe, which generally comes into contact with patient's bodily fluids and tissue, to be removed for chemical cleaning and disinfection and then reattached.

In order to be reusable, the probe was assessed for sterilisation using the Johnson

and Johnson STERRAD 100nx system by Johnson and Johnson, California. The probe underwent extensive bioburden testing to prove adequate elimination of microbes and functional assessment of the fibre optics to confirm functional integrity of the system after multiple rounds of sterilisation.

4.4.4. Coupling of probe to laser and spectrometer

A single 2 m long 200 μm diameter fibre was butt coupled to the pigtailed fibre from the laser to facilitate delivery of the excitation light from the laser to the filter module. A second specially designed fibre bundle guided the tissue emission from the filter module to the spectrograph. The bundle consisted of 110, ultra-low OH, 100 μm diameter fibres packed in a round geometry at the filter module end but spread out into a parabolic arc and coupled directly to the spectrograph at the other end (not all fibres are shown in Figure 26). The round end minimizes the losses when coupling to the flexible probe, and the parabolic end corrects for the image curvature that occurs when using straight slits or vertically aligned fibre arrays as spectrograph objects (Huang et al., 2001), (Zhao et al., 2007). Curvature correction results in superior signal to noise ratio and a higher spectral resolution. That latter was accurately measured using the narrow emission lines from a Mercury/Argon (Hg/Ar) standard lamp. The spectral resolution was weakly Raman shift dependent, reaching a maximum of 7.7 cm^{-1} at 2000 cm^{-1} , (931.2 nm) and varying from 11.7 to 7.8 cm^{-1} with an average of 9.3 cm^{-1} from 395 to 1800 cm^{-1} (810 to 914.3 nm).

4.4.5. Optical filtering of the tissue emission

Multiple stages of optical filtering were used within the system. On the emission collection pathway, long pass (LP) interference filters, with optical densities (OD) ranging from 4-6 for Raman shifts < 395 cm^{-1} ($\leq$ 810 nm) were used. These LP filters attenuated the elastically scattered laser light from the tissue, while allowing $\geq$ 90% transmission of Raman scattered light from 600 to 2000 cm^{-1} (824 to 931 nm). From 395 to 600 cm^{-1} the transmission of the LP filters increased smoothly from $\leq$ 1 to $\geq$ 90%. The role of the LP filters was to reduce measurement noise by stopping significant amounts of elastically scattered laser light from: a) reaching the CCD directly, and b) from generating any emission with Raman shifts from 395 to 2000 cm^{-1} in the optical components of the collection pathway. The 1st stage LP filter (OD4), shown in Figure 26, was hard coated directly onto the incident surface of the collection fibres of the flexible probe (Lightwaves2020 Inc., CA, USA). A 2nd stage LP filter (OD5-6) was a hard coated 12.5 mm diameter substrate (BL01-785, Semrock Inc., NY, USA). The 3rd stage LP filter (OD5-6), shown in Figure 26, was a hard coated 60 mm diameter substrate (Pixel Teq Inc., FL USA) placed in the spectrometer before the grating.

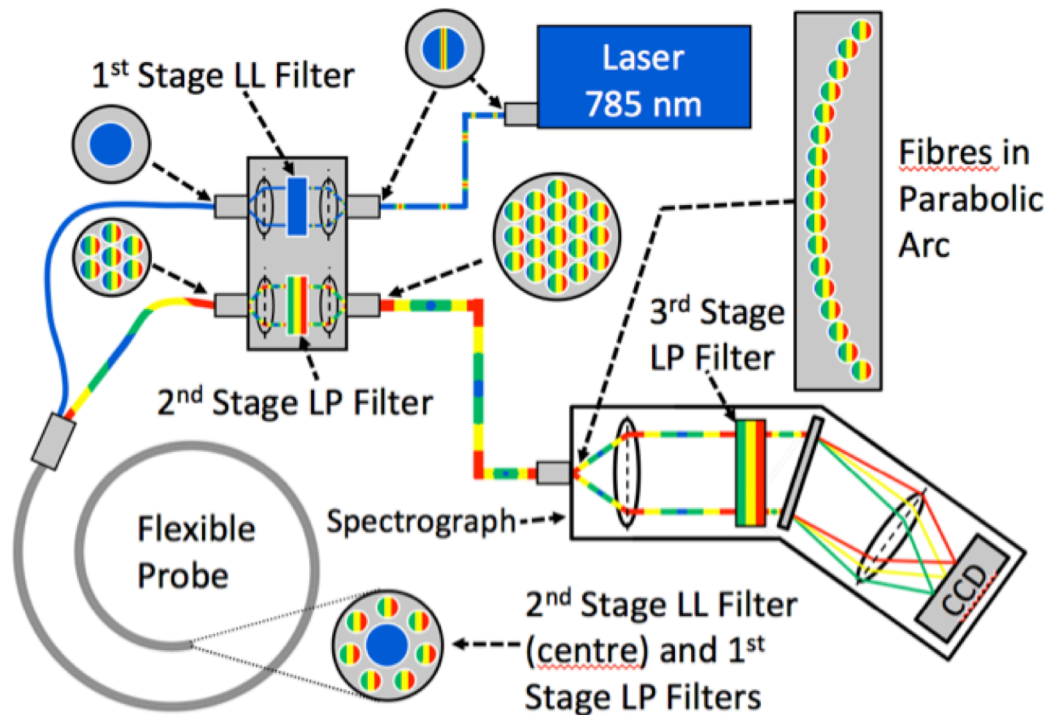


Figure 26. Schematic diagram of the Raman system (not to scale).

Colouring is for illustration purposes, although the wavelength equivalents of the colours used are meant to represent relative changes in wavelength as light propagates through the system. Laser emission (blue colour, 480 nm) < anti-stokes emission (Raman) (green, 510 nm, yellow, 580 nm, and red, 650 nm). Fibre and fibre bundle cross-sections show approximate packing arrangement and relative magnified size, not actual size, and in some cases not all fibres are shown. LL = Laser Line, LP = Long Pass.

4.4.6. Optical filtering of the tissue excitation

On the excitation light pathway, both short pass (SP) and band pass (BP) interference laser line (LL) filters were used. These filters stop significant amounts of: a) non-resonant laser emission, and b) 395 to 2000 cm^{-1} emission, generated in the optical components of the excitation light pathway by the laser light, from reaching the tissue surface. The 1st stage LL filter (shown in Figure 26) was a BP filter hard coated on a 12.5 mm diameter substrate (LL01-785, SemRock Inc, NY, USA). This filter has a full width at half maximum (FWHM) pass band of 48 cm^{-1} (3 nm), with $\geq 90\%$ peak transmission at 785 nm and OD 5-6 attenuation for Raman shifts $\geq 129 \text{ cm}^{-1}$ (793 nm). The 2nd stage LL filter was a SP filter hard coated directly onto the exit surface of the excitation fibre of the flexible probe (Lightwaves2020 Inc, CA, USA). The SP filter has a transmission cut off at a Raman shift of 239 cm^{-1} (800 nm) allowing $\geq 90\%$ transmission of the laser line at 785 nm, but attenuation (OD4) of all Raman shifts from 395 to 2000 cm^{-1} (810 to 931 nm).

4.4.7. System control

Control of the system was implemented by a personal computer (PC) using a custom designed program. A footswitch was used by the clinician to trigger data acquisition. Once triggered, two consecutive spectra were obtained at a pre-set exposure time; the first was a background spectrum and the second was taken at full laser power, (preset up to a maximum of 150 mW). The program then subtracted the background spectrum from the spectrum obtained under full laser power. The difference spectrum was calibrated for the response of the system (see Figure 41), smoothed and the auto-fluorescence background removed by an iterative polynomial fit (Zhao et al., 2007) all in real time. The resulting Raman spectrum was displayed on a large light emitting diode (LED) type monitor within 3 seconds for a pre-set exposure time of 1 second. All spectra were then automatically saved to the hard drive.

4.4.8. System packaging

The complete system: spectrometer, laser, PC, and associated electronics were housed in a single enclosure and mounted on a moveable cart as shown in Photograph 3. A small LCD touch screen was provided for program changes (calibration, exposure time, power settings, etc.). The large LED monitor screen was used to display the captured spectra. The filter module and bifurcated probe were suspended on an articulated arm which was approximately 1.5 m long fully extended. The arm also had 270° of rotational freedom around its base mounting spigot which increased the system's positioning flexibility within the surgical suite. The articulated arm served four purposes, a) it reduced the clutter (equipment and personnel) around the patient during the procedure, b) it allowed for a shorter flexible probe to be used which reduced the laser induced emission in the fibre which may not be completely blocked by the optical filters located at the tip of the flexible probe, c) it reduced potential to compromise the sterile surgical environment, and d) it reduced the potential for snagging the probe on other equipment.



Photograph 3. Photograph showing the complete system in use mounted on a mobile cart.

The Raman system is being used and is in between the Sonowand neuronavigation system and the operating microscope.

4.4.9. Summary

The Raman *in vivo* system described here represents an effective, safe system consisting of a sterilisable fibre optic probe and a unique set of filters to enable high quality Raman spectra to be obtained. The system is high performance and easy to control using a footswitch. The use of this system during surgery as part of an *in vivo* protocol will be detailed in the next chapter.

4.5. Raman Surgical Protocol

4.5.1. Introduction

This chapter deals with study design, *in vivo* data collection methodology with particular attention to specimen annotation with MRI and histology and quality control. The aim is to provide a framework to enable future *in vivo* applications of Raman spectroscopy in brain surgery.

4.5.2. Study design

The Raman study was an *in vivo* pilot study non-invasively obtaining Raman spectra from 40 brain tumours during surgery. The subjects are those fulfilling the inclusion criteria of all consenting adults undergoing brain tumour surgery at Charing Cross Hospital. The proposed duration is 18 months.

The study pathway is outlined in Figure 27. The research participant will be from a pool of brain tumour patients referred via GPs and hospital specialists into our neuro-oncology system and the case will be discussed in a multidisciplinary meeting run in collaboration with radiologists, oncologists and neurosurgeons. A clinical decision to operate in order to diagnose the tumour prior to further treatment, improve symptoms or resect the tumour to improve survival will be made here as part of the normal NHS Neuro-oncology patient pathway. The patient will be seen within the same week in our Neuro-oncology clinic where the possible diagnosis and treatment strategy will be explained, and consent for surgery/research will be undertaken by the principle investigator and nurse specialist. Patients selected for the study will have full mental capacity to take part in research and will not be coerced into doing so. They will have surgically easily accessible tumours which will be relatively superficial to enable data to be collected easily. Within a week the patient will have planning and spectroscopy MRI scans and will then be admitted for tumour biopsy or removal. Consent will be reaffirmed prior to surgery. The operation will be conducted as normal and the study will not require any additional surgery or change to surgical technique. During surgery, data will be collected as to the wavelength of scattered laser light by shining the probe laser into the surgical cavity. The centre of the tumour cavity and the edge of the tumour as well as unaffected normal brain will be analysed in a matter of seconds using a non-contact technique. Multiple readings (it is estimated 8 per patient) will be taken depending on the quality of the spectra gathered. The operation will proceed to the extent of resection normally undertaken using current methods. The data of the first 40 patients thus operated will be analysed to see if the Raman spectra of tumour tissue is distinct from that of normal brain tissue. This analysis will be done using industry standard Windows based Raman spectroscopy software. These spectra will then be compared to independent gold standard histopathological classification which would be done routinely as part of patient care in the patient pathway. No human tissue will be taken for analysis at any time. This study is being

run concurrently with the iKnife mass spectral analysis of tumours in the same patients and results will be compared for cross-validation of both Raman and Mass Spectral technologies.

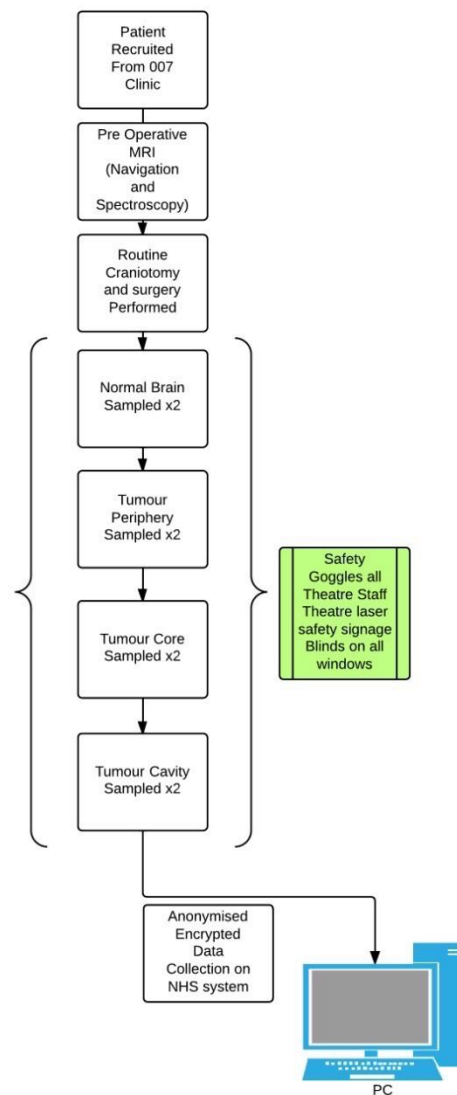


Figure 27. Raman study pathway diagram.

The overall study workflow, from patient recruitment to data analysis.

4.5.3. Methods

The steps in data collection that I used during surgery are represented in Figure 28.

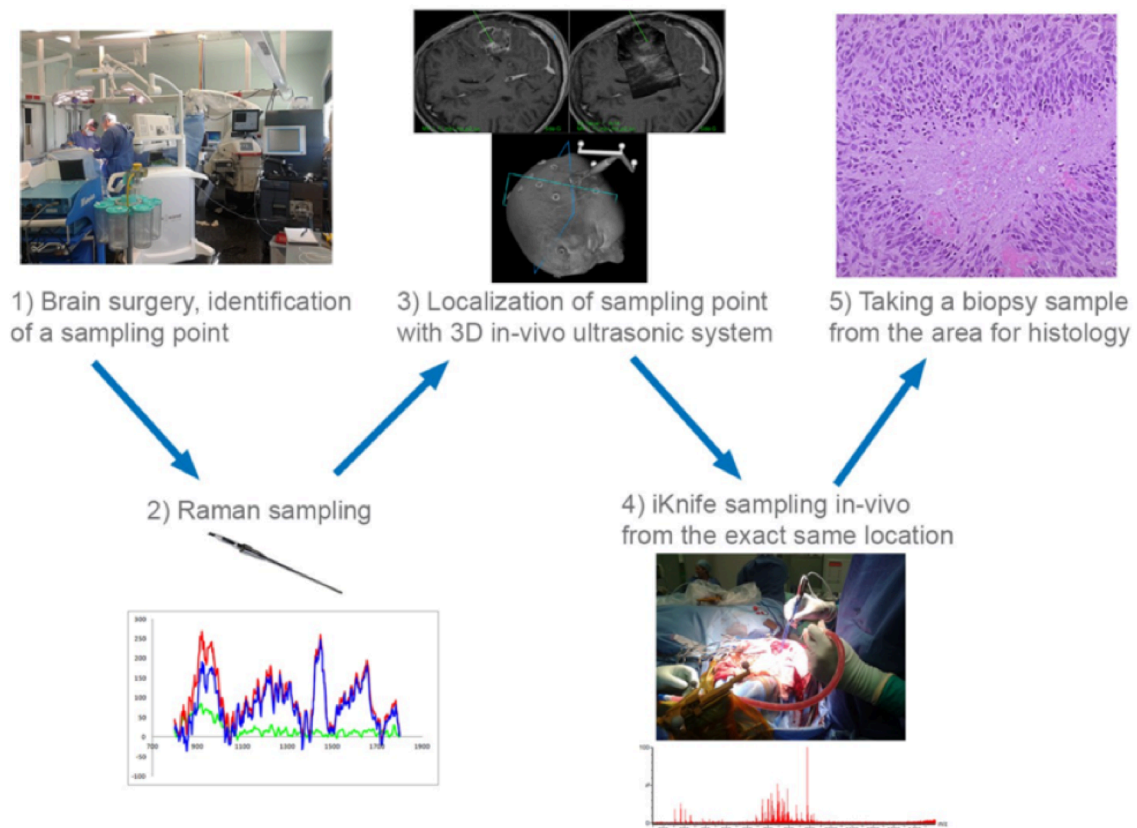


Figure 28. Flow chart of ex vivo sampling in the operating theatre.

The samples were divided into sections and multiple spectra were obtained with the same precautions for ambient background light.

Preoperative Preparation

- Inserting the optical filtering lens to the microscope and drape microscope
- Switch Raman system on and wait for Windows to bootup and CCD to cool.
- Ensure Raman system clock and Sonowand clock are synchronised

During Surgery

- Attach the probe to the wand
- Ensuring the room is dark
- Wash the spectral area with saline before and after acquisition

- Take a reading in accordance with guidelines (see section 4.2.2.10)
- Moving the light screens away
- Taking screenshots: correlation using time and location
- Name each location anatomically
- Analyse reading; repeat if necessary
- Move on to the next area; repeat the above
- Clean the probe tip with saline and lens wipe after, then move on to *ex vivo*
- *Ex vivo*: take tumour - divide into sections - place on glass slide - take reading from slide first - then take reading from a tumour - try to maintain equal distance between the sample each time - remove sample and replace with another on glass slide - ensure room is dark and screen is turned away - take multiple readings.

4.5.4. Matching spectra to MRI and tissue samples

- The Raman system has an internal timer, which I matched to the time on the Sonowand to the exact second
- The number of readings I took were noted down by IP at any one point
- The number of readings were verified manually by myself
- Screenshots were taken of the location of the acquisition
- I gave a descriptor to each of the locations a spectrum was taken from as I was operating
- These screenshots were later matched to the relevant Raman spectra using the timestamps and verified independently by myself (UF)
- The times on the Raman system and Sonowand were checked multiple times throughout the operation to confirm that they were in sync
- Each Raman spectrum was thus matched to a location on the brain

Check that each Raman spectrum now has

- Time stamp
- Location Descriptor

- Screenshot of location

Tissue Handling

- The samples of brain tissue taken are now passed to the scrub nurse.
- The scrub nurse labelled each sample according to the descriptor stated by the surgeon and a number (to indicate which order they were taken out in)
- These descriptors and the numbers were matched with data noted down by two independent assistants (UF and IP) respectively
- Each sample was then subject to *ex vivo* data acquisition, being careful to sample the tissue in order it was removed

Each sample now has *in vivo* and *ex vivo* spectra. The scrub nurse had clearly labelled the bottles with numbers which are the basis for matching the sample to the histology report post-op. I checked concordance of labelling prior to sending all samples.

4.5.5. Quality control

1. The spectra were taken in a room which was completely dark
2. The only light source was the microscope; to which a filter was attached
3. Any other necessary light sources; the Sonowand screen were turned 180 degrees away from the site of spectral acquisition, before taking a spectrum and during the process of taking them
4. The quality of the Raman spectra I obtained were verified by IP in real time
5. If a spectrum was not of suitable quality, I repeated the reading immediately without moving the probe from the location on the brain.

4.5.6. *in vivo* brain measurement

The system takes two consecutive spectra for one press of the foot switch. The first spectrum is taken at very low power and the second at high power. Both spectra are saved to the C:\Data folder. If the foot switch is held down the system will take multiple low power spectra, and one high power spectrum upon release of the foot switch. All spectra are saved. A low power spectrum acts as a background spectrum

taking into account the any ambient light contamination. This can be subtracted from the high power spectrum to produce a spectrum largely free of background light. However, one should take all necessary steps to reduce the background light as much as possible to improve the quality of the data. The setup I used for *in vivo* data analysis was as follows:

- 1) Position system in the operating theatre at the direction of the clinician. The distance to all measurement locations must be within 2m. Also make sure there are mains power outlets with 2m otherwise an extension cord will be required.
- 2) Plug system and computer monitor into mains supply. Turn on the monitor and press black button to start Raman system. Continue set up while system cools down to -20 C.
- 3) Remove brass caps off of the output of filter connectors and store for later use.
- 4) While the clinician is supporting the sterile probe, remove the black protective end plugs off the 3 legs of the proximal end of the probe and attach the legs to the filter connectors. Leg labelled ILL2 (yellow) to the filter output labelled green. Leg ILL1 (yellow) to the filter output labelled yellow. Leg COL (white) to the filter output labelled white. Be sure to attach the 3 Legs in the right way and be careful not to touch any other part of the probe. Note: leg ILL2 is not used in this application but connecting the leg to filter is recommended to prevent damage to probe.
- 5) Ensure that the clinician removes the blue protective cap off the distal end of probe before use. Keep this cap and the caps from the proximal ends.
- 6) Once system has cooled down a key pad will appear on the touch screen. Enter code 1234 and enter.
- 7) Then advance through a series of windows (not used in this application) by pressing the "continue" arrow on the touch screen. Stop when you reach a black screen with the words "READY FOR SCAN ". Note: you will have to repeat this step every 30 mins as the system will reset after this period of use.
- 8) Check time on system clock (lower right of monitor) is synchronized with clock on surgery guidance system. Adjust system clock if off by more than 5 seconds.
- 9) Position foot-switch in a convenient location for the clinicians.
- 10) Before measurements start, ensure all window blinds are closed both for external and internal windows. Also ensure all fluorescent lighting and non essential LCD monitors are turned off. Ensure all personnel are wearing laser safety goggles.
- 11) During measurements direct IR emission and the computer monitor of the surgery guidance system away from surgical location. Similarly turn the overhead LED lights away from the surgical location. Reposition after each measurement is complete.
- 12) After all measurements are complete retrieve probe from the clinician, but do not disconnect from system. Ensure tip end is clean by wiping with saline soak gauze. Attach spacer to the distal end of probe end so that the end of probe is flush with exit surface of spacer. Tighten locking screw to just gently hold probe in position. **DO NOT OVER TIGHTEN OTHERWISE PROBE WILL BE DAMAGED.** Place spacer into the two reference standards and take 20 spectra, 10 for each standard. Note: hold spacer down so there is no movement of the probe during these measurements.
- 13) Turn on fluorescent lamps and aim probe towards lamps. Make sure no person or object is blocking the light path. Take one spectrum. If touch screen indicates measurement is "spectrum is saturated" repeat measurement with probe slightly pointing away from fluorescent lamp. Repeat positioning until a spectrum is obtained.

- 14) Protective laser eyewear from personnel can now be removed, and lighting returned to normal.
- 15) Remove spacer. Disconnect probe from system and reattach the 4 protective caps (3 black and 1 blue). Place probe and spacer in tray for sterile reprocessing. Replace brass caps on output ends of filter connectors.
- 16) Data from the current patient is stored in C:\data folder. Make a new folder in the C:\ drive; name this folder with the current patient ID (consult clinician on appropriate patient ID to use). Move all data of current patient (including calibration spectra) to the newly created folder. Plug in a USB flash drive and copy the newly created folder and contents. Consult clinician about ethical use and storage of data backup.
- 17) Remove USB flash drive and press black button to shut down. System will automatically shut down once it has warmed up. Do not cause a premature shut down by disconnecting from mains power unless there is an emergency.
- 18) If system has been soiled in any way it must be wiped down and disinfected before leaving operating room.

4.5.7. Histopathological analysis

All biopsies I collected as part of the study protocol were submitted and processed at the Department of Histopathology, Imperial College Healthcare NHS Trust, London. Tissue samples underwent standard fixation in formalin and paraffin embedding. Tissue samples obtained from this study were initially sent immediately as fresh samples to the pathology laboratory by hand.

The specimens were immediately frozen and were then fixed as required. 4 µm thick paraffin sections were taken for haematoxylin-eosin staining and immunohistochemistry, and 10 µm thick sections for DNA isolation. Immunohistochemistry was performed according to manufacturer instructions.

The following primary antibodies were used: mouse anti-mutant IDH1-R132H (1:150, DIAH09, Dianova), rabbit anti-ATRX (1:750, HPA001906, Sigma), mouse anti-MGMT (1:500, MAB16200, Chemicon International Inc.). In a subset of cases IDH1 and IDH2 sequencing and/or MS-HRM PCR assay for MGMT promotor methylation were carried out at the Molecular Neuropathology Laboratory of the Division of Neuropathology, Institute of Neurology Queen Square, University College of London.

4.5.8. Conclusion

An *in vivo* surgical protocol for the successful acquisition of Raman spectra during brain surgery has been described. Steps taken to ensure quality control of data and minimise data loss through sample labelling errors have been outlined. A solution to the problem of matching Raman data accurately with existing imaging datasets and histopathological reports has been proposed.

The next step is successfully analysing the data produced to answer key clinical and research questions. A reproducible data analysis methodology will be presented in the following chapter.

4.6. Data Analysis Methodology

4.6.1. Introduction

The previous chapters have dealt with system setup, method development and the protocol for using Raman spectroscopy during surgery for the first time in humans. The analysis of the large volumes of complex data that will be produced is a vital next step in the exploring this modality as a platform for tissue detection during surgery. This chapter will introduce the basics of multivariate data analysis in terms of dimensionality reduction, model building and measuring classifier performance. The concept of data analysis as a series of data experiments will be discussed and the basis for the structure of the results chapters will be made clear.

The aim is to develop a data methodology which can be used as a template for future *in vivo* Raman systems and recognise issues critical to future data driven diagnostic trials on rare tissue such as sample size and number of tissue samples required to train accurate predictive models.

4.6.2. General principles

The Raman dataset represents a matrix of wavenumbers and relative intensities. Feature extraction of this complex dataset can be solved using a similar workflow to face recognition problem solving (Sahoolizadeh et al., 2008) and other pattern recognition tasks as the underlying principles are the same, namely:

- Data pre-processing
- Dimensionality Reduction: PCA, LDA
- Feature Extraction
- Classification: using neural networks, virtual machines (multivariate statistical models), ensemble methods
- Cross Validation to assess model performance

The data collected includes the raw spectroscopy data as well as metadata - meaning all other data pertaining to the patient or tissue samples such as diagnosis, patient age and gender etc. The data collected must have the correct metadata associations which must be checked throughout the data processing procedure to ensure data fidelity.

One possible solution to process the data for the stages above is as follows:

- Data selection, including predictor selection, label selection
 - Need to organise spectroscopic data and metadata, format data, label data and determine relationships
- Machine learning model application

- Need to select machine learning models to train
- Model performance assessment
 - Common methods for assessing performance: Confusion Matrix, ROC curve
- Exporting model for use on new data
 - Programming environment for classification e.g. Matlab or R code

The approach I used is iterative and is summarised below (Baker et al., 2018).

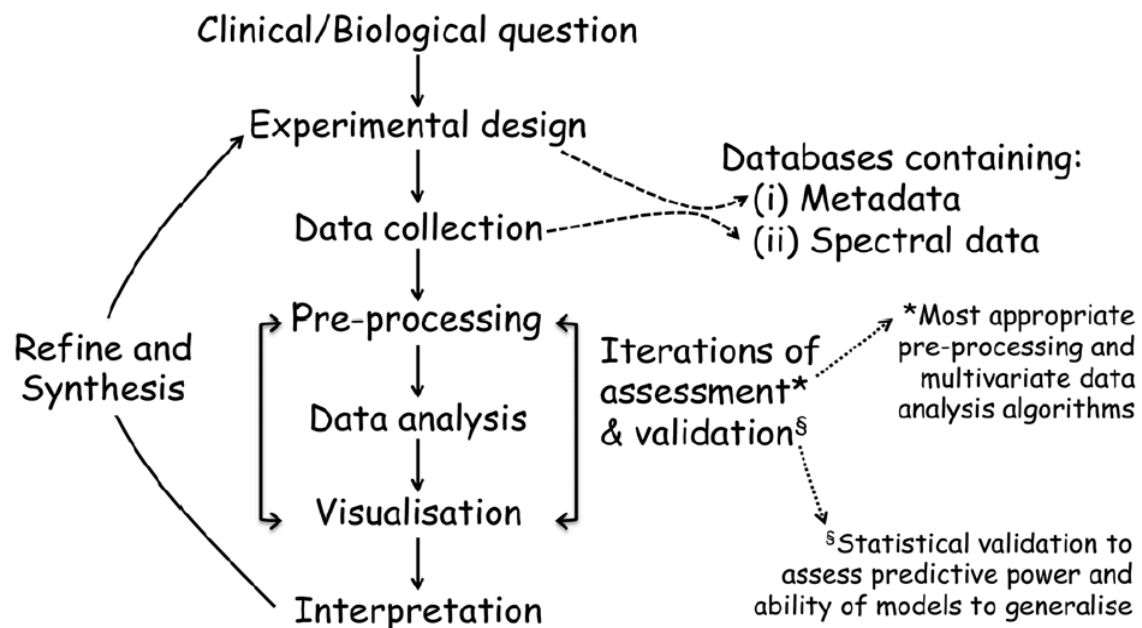


Figure 29. Data flow paradigm in spectroscopy, taken from Baker et al. (Baker et al., 2018)

This emphasises the data analysis as an experiment in itself, with analysis seeking to answer specific clinical questions and then the best methods of modelling the data and predicting outcomes are investigated and selected. This may inform further types of data labelling and categorisation which may lead to more improved models being generated depending on which relationships are revealed during data analysis. The functions of data analysis include:

- Data exploration: data compared to find out relationships based on variance (PCA)
- Classification of data: model built on training dataset and then validated on an unseen testing dataset.
- Explanatory analysis: loadings vectors, feature selection

The different data analysis approaches and related machine learning processes are summarised in Table 6.

Data analysis approach	Research question being asked	Algorithm features	Chemometric MVA solutions
Exploratory data analysis	• What can we find out about the relationship between these samples? • <i>e.g.</i>, to investigate gender or age	• No prior information is required • Unsupervised method is used	• Principal component analysis (PCA) • Dendrograms • Self-organising Map (SOM) • Linear discriminant analysis (LDA)
Classification of samples	• Can I classify samples into different groups? • <i>e.g.</i>, case-control studies	• Prediction of unseen samples into groups • Supervised learning is used • Should include validation	• Partial least squares-DA (PLS-DA) • Support vector machines (SVM) • Random forests (RFs) • Neural networks • PLS regression
Quantitative analysis	• Can spectra be correlated with a continuous variable • <i>e.g.</i>, time, age, disease status, drug dose	• Allows quantitative prediction of unseen samples to some continuous variable	• Principal component regression (PCR) • Support vector regression (SVR)
Explanatory analysis	• What are the differences in chemical or physical state between samples • <i>e.g.</i>, can I find vibrational features that relate to group separation or a continuous feature	• Highlights spectral changes as function of group membership • Need to know which group each spectrum belongs to	• Many algorithms can be used: • Based on loadings vectors; <i>e.g.</i>, LDA, PLS-DA, PLSR, PCR • Feature selection in RFs; use of Gini • Genetic algorithms (GAs) as input selectors for MVA

Table 6. Comparing machine learning methods for chemometrics in Neurosurgery. Taken from Senders et al. (Senders et al., 2018)

Therefore, only specific algorithms and analysis approaches will be presented dependent on the clinical question asked for each results section.

4.6.2.1. Signal pre-processing

The signal pre-processing stage I used involves converting the raw data into a normalised, fluorescence subtracted dataset to allowing meaningful comparisons between spectra. Fluorescence within the near infrared range causes interference with Raman signal, and in many cases the Raman signal rides on top of background environmental fluorescence and fluorescence from the optical fibre being used for data collection. I used a polynomial subtraction to subtract background fluorescence as described in section 3.3.2. Pre-processing aims to make all of the data comparable on the same scale and consists of the following steps:

- Fluorescence removal
- Correction for fibre fluorescence based on background readings taken in darkness at the end of the previous surgical case
- Data normalisation using AUC in Microsoft Excel.

The specific steps used for Raman signal pre-processing were as follows.

To pretreat the data, I imported the saved spectral Raman shifts and intensity values

into a spreadsheet along with the calibration standard measurements obtained for each patient. The latter were compared to the preset system calibration values and if necessary adjustments were made to that patient's data to account for any slight differences. For the analyses a more advanced iterative 5 point quadratic Savitzky–Golay (Savitzky and Golay, 1964) smoothing was chosen to remove high frequency spectral noise from the calibrated raw data. This was the highest level of smoothing that did not affect the intensity of the narrowest Raman peak (the phenylalanine peak at 1003cm⁻¹) observed in some brain tissue spectra. The corrected raw spectral signal contained the Raman signal superimposed on the tissue autofluorescence background signal. An iterative polynomial background subtraction algorithm (Zhao et al., 2016) was used to remove this background autofluorescence. The Raman spectrum for each data site was normalised using the Area Under the Curve (AUC).

For tentative identification of molecular changes, I averaged the Raman spectra by IDH, ATRX and MGMT groups and compared with reference spectra from biomolecular standards. Two main standards were used. The first was a 2-hydroxyglutarate (2-HG) a metabolite which is known to accumulate in IDH mutated cells (Dang et al., 2009). A pure sample of 2-HG was obtained from Sigma-Aldrich Canada Co. (Part # 90790, Oakville, Ontario, Canada) and measured with a sister Raman system to the one described here. The second reference sample was human DNA (deoxyribonucleic acid). This sample was also obtained from Sigma-Aldrich Canada Co. (Part # D4642) but was measured with a different system (Short et al., 2006). That system was accurately calibrated allowing for an accurate comparison of relative peak intensities and positions with those measured with the current system.

The data was now considered fluorescence subtracted, standardised and normalised ready for the next step in analysis.

4.6.2.2. Dimensionality reduction

One of the difficulties inherent in multivariate statistics is the problem of visualising data that has many variables. The MATLAB® function [plot](#) displays a graph of the relationship between two variables. The [plot3](#) and [surf](#) commands display different three-dimensional views. But when there are more than three variables, it is more difficult to visualize their relationships.

Complicated systems with many variables can be difficult to summarise and generalise. However, mathematically complex systems often have a small number of core governing principles interacting in a non-linear fashion to produce what often looks like highly complex behaviour. When measuring hundreds or thousands of variables it may be possible to identify related variables through these deeper system principles and create a smaller number of variables which govern the whole system. PCA achieves this through generating core variables, called principle components, which can explain the variance as a linear combination of original variables. Keeping the principle components orthogonal eliminates redundancy and this produces a new 3D plot to represent all of the data and the variance contained within it when 3 principle components are used. In reality there may be hundreds of

principle components, resulting in a multidimensional hyperplane. However, often just the first few principle components contain enough information to explain 90% of the variance and going beyond 10 or 20 principle components to abbreviate the data is unnecessary. I used the Matlab function [pca](#) to compute the principal components.

It is worth noting that using PCA does not require information regarding the identity of the classes of data being analysed. It is an unsupervised technique designed simply to identify those principle components which can separate the raw data as much as possible.

4.6.2.3. Classification using machine learning

Learning can be defined as the process of using past experience or data to create a model to predict future behaviour given an unseen experience or set of data. Machine learning is the ability of mathematical algorithms to perform this learning process without being explicitly programmed to do so. In general terms learning can be thought of as its simplest form as a classification task, with the algorithm attempting to successfully classify new data according to predefined data labels. It can be considered as having correctly classified the new data if the assigned data label matches the true data label (often called the 'ground truth'). The learning algorithm is therefore constantly attempting to classify new data relative to the ground truth.

Algorithms used in machine learning employ a number of different approaches. Attempts can be made to find out if the data can be made to cluster in groups using clustering algorithms such as K-Means. Simple regression may define different data groups, or simple dimensionality reduction using PCA. Decision trees can be used to iteratively use rules to separate data. Ensemble methods can combine different algorithms to improve performance. Neural networks can process data in many processing layers in an attempt to separate it into correct classes, and support vector machines modelling which aims to find a hyperplane with the maximal separation between data groups. A summary of different machine learning algorithms is shown in Figure 30.



Figure 30. Machine learning methods grouped into methodological classes.

Machine learning can be classified into 2 broad groups: supervised and unsupervised learning. Unsupervised learning occurs when the 'ground truth' or class labels are unknown, and relationships between unstructured data are required. Supervised learning is used by the vast majority of machine learning methods. It consists of allowing the algorithm to know the identity of the classes in a training group of data. The algorithm is instructed to map the input variable (in this case Raman spectra) to the output variable, which is the class identity in the most accurate way possible. The performance of this model is then validated against an unseen testing dataset where the results of the algorithm are compared with the true class values. This approach uses 2 datasets: a training dataset and a testing dataset. The original data can be split in a number of ways with the commonest split being 80% training and 20% testing, but 70/30 and 50/50 splits can be used to estimate performance.

For multivariate statistical analyses I pretreated data for all spectra and imported it into Matlab (MathWorks, MA, USA), Statistica (TIBCO, CA, USA) and Rattle (Togaware Pty, Canberra, Australia). I used a range of supervised and unsupervised machine learning techniques for model development including Principle Component Analysis (PCA), Linear Discriminant Analysis (LDA), Random Forrest (RF), Boosted Trees (BT) and Support Vector Machines (SVM) as well as Partial Least Squares/Linear Discriminant Analysis (PLS-LDA). My code used for the analyses will be introduced below.

4.6.2.4. Cross validation

Cross validation is a model validation technique to assess the predictive power of the model in practice on an independent dataset. There are 2 classes of cross validation: exhaustive and non-exhaustive. Exhaustive cross validation was used in this instance. The commonest form is Leave One Out Cross Validation (LOOCV). In a data set consisting of n observations, one observation is left out; a training model is built with the remainder of observations and then an attempt is made to classify the observation that has been left out. This is repeated for $n-1$ observations, and the predictive power is then calculated. Non-exhaustive cross validation techniques include k fold cross validation. In this technique the data is randomly divided into k equal sized subgroups. Of the k subsamples, a single subsample is retained as the validation data for testing the model and the remaining $k-1$ subsamples are used as training data.

The LOOCV method that I used builds a classification model on $n-1$ training spectra, where the remaining spectrum is used as the test spectrum. This procedure is repeated n times such that each spectrum is left out once. The results are tabulated as a confusion or error matrix from which sensitivity and specificity can be calculated.

4.6.3. Dimensionality Reduction: PCA and LDA

I performed PCA analysis in Matlab R2017b. The process was as follows.

Pre-processed data was imported as an Excel file. This file had two regions of interest: the spectral data (numerical) and the spectral identifiers (string of text). The identifiers contain a text string containing the spectrum identification number, histological diagnosis, biopsy status (yes/no), genetics results from the biopsy (all WHO 2016 recommended genetics arrays) as well as *in vivo* or *ex vivo* designation. The imported excel file had Normal and Tumour classes imported as strings (allowing Matlab to analyse the cells correctly). Importing without this function leads to erroneous labelling of some cells as non-string entities which prevents PCA analysis. PCA analysis is then performed on the data. I created the following Matlab script to automatically imports data, labels Normal and Tumour as strings after searching from them and replacing with a numerical number and then performs PCA analysis with a 3D plot as an output using gscatter:

```
%This script will import Raman data and prepare it for PCA.
path= '/Users/babarvaqas/Documents/MATLAB/Raman.xlsx'
load spectra
[a,b,~]=xlsread(path);
c1=strfind(b(4,:), 'Tumour');
C2=strfind(b(4,:), 'Normal');
lab = cell(numel(c1),1);
for n = 1:numel(c1)
```

```

    if isempty(c1{n})
        lab{n,1} = 'Normal';
    else
        lab{n,1} = 'Tumour';
    end
end
%Problem is lab has one extra column in the beginning- this needs to be
%deleted manually
type=lab';
[~,score,pcvars]=princomp(spectra');
x=score(:,1);
y=score(:,2);
z=score(:,3);
gscatter3(x,y,z,type,{'b','g'},{'.','.'},15);

```

It is important to load spectra as observations in columns, define as a numeric matrix and name as 'spectra'. Spectral labels must be imported as a single column defined as a numeric matrix and be labelled as 'lab'. The function gscatter3 is available as a downloadable app via Matlab and enables a 3D scatter plot output of the principle components. This is necessary as the built in 'scatterplot' function in Matlab only allows 2D scatterplots. If the class being allocated to is a string, it must be replaced with a number for gscatter to work as it only recognises a numeric matrix, not a string of characters. To do this the [unique](#) function is used. For cases with more than 1 label being compared, I used [unique](#) to go through the lab array and assign a unique ID to each string. This array, called 'id' was used to separate out the data stored in spectra for each category so I could plot these on a graph with the different colour and marker:

```
[~,~,id] = unique(lab);
```

It is then possible to successfully use gscatter to construct a 3D plot:

```
gscatter3(x,y,z,id,{'b','g','r','y'},{'.','.', '.', '.'},15);
```

In the example above the scatter plot will consist of a 3D axis (x,y,z) with 4 classes, each with its own unique colour (blue, green, red, yellow) based on the class label in the 'id' matrix and will be represented as small spheres '.' An example of the output is shown in Figure 31.

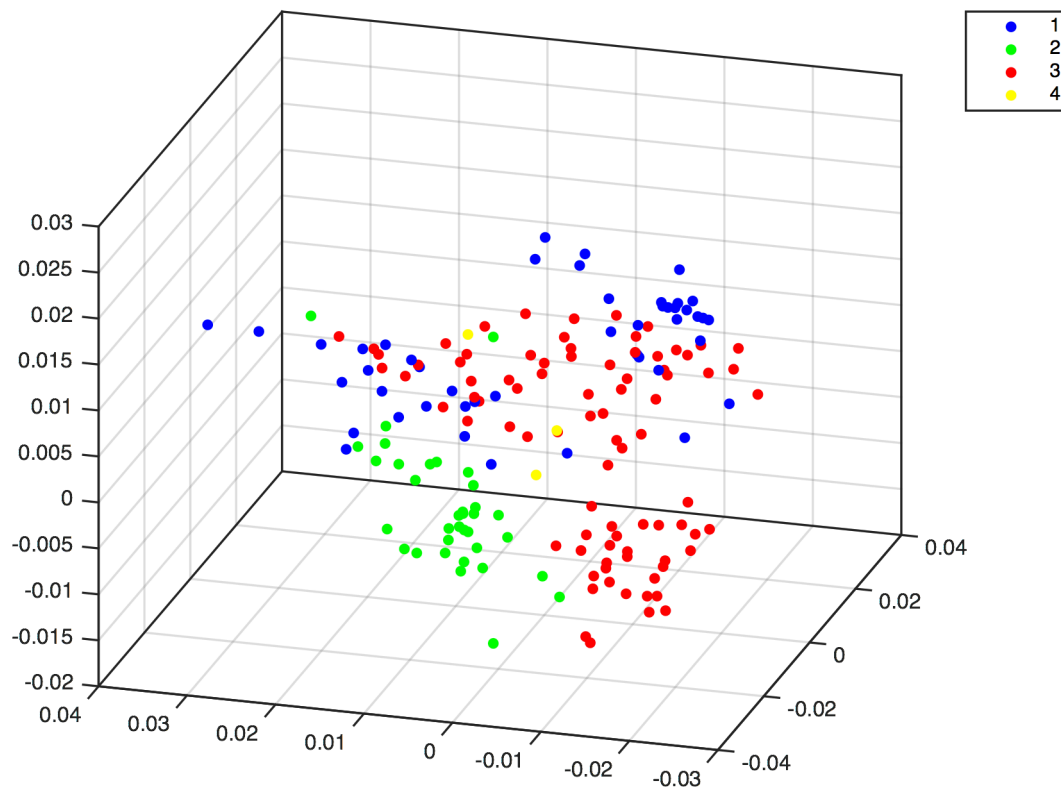


Figure 31. PCA plot using gscatter showing breakdown by tumour type.
3D plot comparing GBM (group 3) astrocytoma (group 1), oligodendrioglioma (group 4) and choriocarcinoma (group 2).

This script is extremely flexible and allows for a large number of class identities to be plotted, for example the script below is used for further breakdown of the same data in Figure 31 to look at the effect of tumour grade and tumour type:

```
[~,~,id] = unique(lab);
[~,score,pcvars]=princomp(spectra');
x=score(:,1);
y=score(:,2);
z=score(:,3);
gscatter3(x,y,z,id,{'b','g','r','y','m','c','w','k'},
{'.','.', '.', '.', '.', '.', '.', '.', '.'},15);
```

Figure 32 shows the output of this function: I generated a good quality 3D PCA plot.

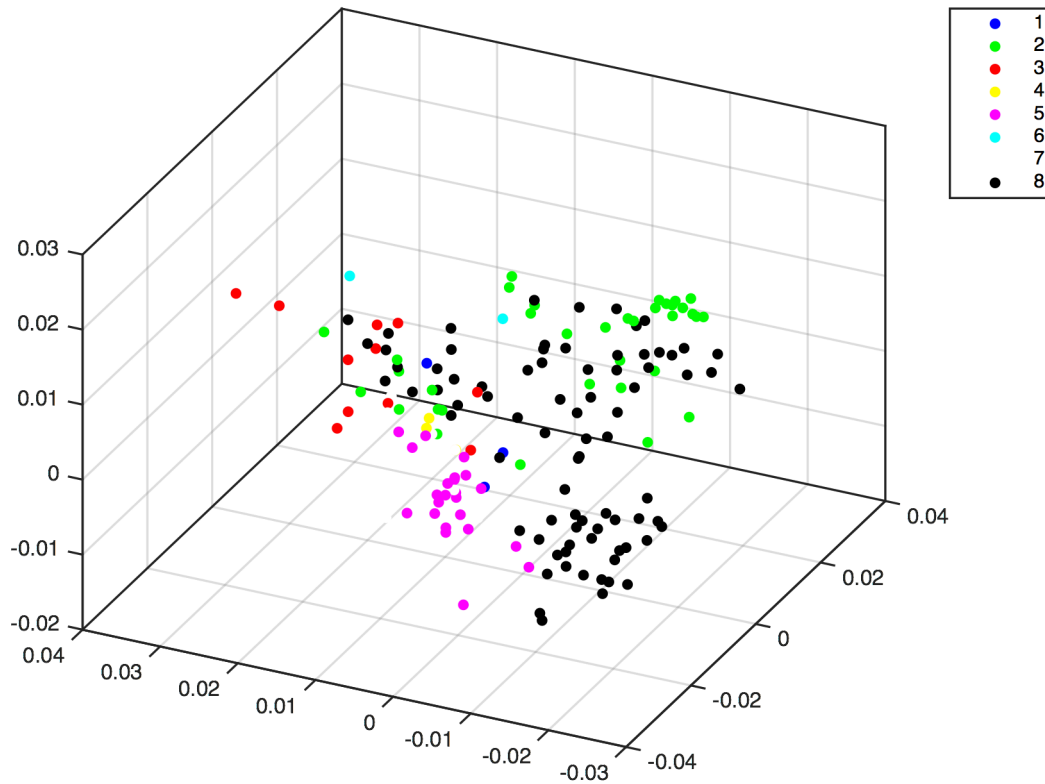


Figure 32. PCA plot using gscatter showing breakdown by tumour grade and type.

Class labels include GBM (8), G2 astrocytoma (2), Choriocarcinoma (1), G3 oligodendriglioma (6), G2 oligodendriglioma (3), G2 Oligo/GBM (4), G3 Oligo (GBM) (7), G3 astrocytoma (5).

I built LDA models as follows using the 'fitdiscr' function in Matlab R2017b. Once labels had been loaded (although not used in the computation they were required for labelling the PCA plots) and LDA model was built as a one step function:

```
Mdl = fitcdiscr([Raman source data],[Labels]);
```

This model can then be used as a predictive tool by instructing Matlab to predict on new data using this model as the basis:

```
label=predict(Mdl,[Raman data matrix to be tested]);
```

The output is the label best matching the Raman data matrix which is being tested.

4.6.4. Support Vector Machine Learning

Support Vector Machines (SVMs) represent a powerful technique for non-linear classification problems. The optimal separating hyperplane between observations is sought by maximising the margin between the closest points between the classes.

The data points on this boundary are called support vectors, and the middle of the margin between the 2 classes is the hyperplane as shown in Figure 33. This technique therefore looks at the margins between the 2 classes of data and seeks to find the plane that best separates them. The margin is computationally processed so that those points on the wrong side of the margin are weighted negatively, with non-linear separation margins being computed using kernel techniques. The Gaussian radial basis kernel was used to compute higher dimensional hyperplanes enabling a non-linear margin to be computed in an efficient manner using the 'kernel trick' of using the inner products in a higher dimensional space rather than the actual data points themselves.

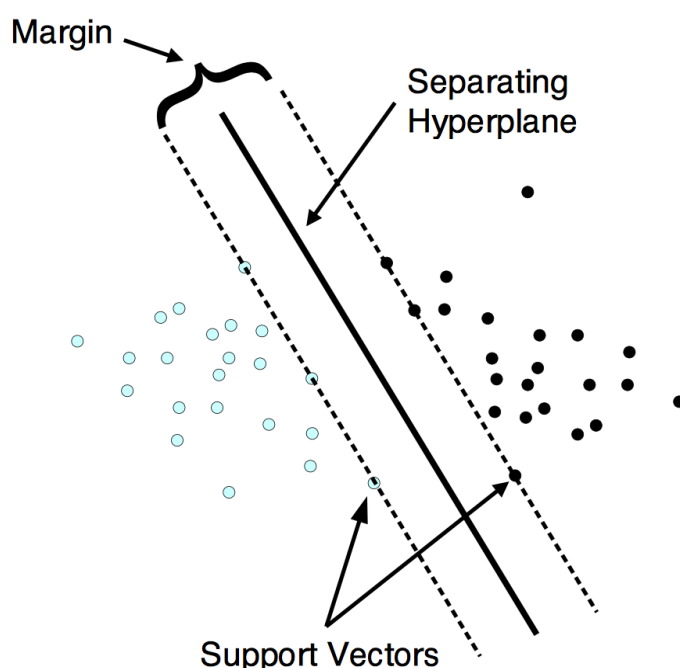


Figure 33. Classification using SVM: 2 classes separated by a linear margin.

I first converted pre-processed and labelled data to .csv format to enable importing into the 'R' statistical package. Rattle (Togaware Pty, Canberra, Australia) was then used to load the data into a graphical user interface for further manipulation. The Rattle package enabled fast calculation of SVM using the e1071 implementation of Chih-Chung Chang and Chih-Jen Lin's 'libsvm' and supplements this with automatic confusion matrix and plotting of ROC curves.

Parameters were set for training and testing dataset split (the most frequently used setting was 70/30). The data is then processed as variables or labels. Variables are the Raman wavenumbers. The labels refer to class labels, with the particular feature being investigated selected as the target variable. The SVM model was then trained on 70% of the data chosen at random and then tested on 30% of the original data used as a validation dataset.

4.6.5. Ensemble methods

Decision trees consist of a like model of decisions with branched consequences. In classification problems they represent classification decisions based on measured parameters and relate to the final classification output. At each branch a decision tree rule is followed - for example, whether a reading is above or below a particular parameter value. Sequences of trees can be used to make a final decision, processing parameters in various sequence combinations.

Decision trees can be combined in different ways as part of ensemble methods such as Boosted Trees (BT) and Random Forrest (RF). In the RF model random decision trees are combined with bootstrap aggregating (bagging) to increase precision. Each model in the ensemble is trained on a randomly chosen subset of decision trees from the training set, with each model in the ensemble being given equal weight. BT involves building an ensemble by emphasising previous failed training instances to boost model performance.

Here, I used the MATLAB function TreeBagger to perform an analysis similar to Random Forest. TreeBagger grows the decision trees in the ensemble using bootstrap samples of the data. Also, TreeBagger selects a random subset of predictors to use at each decision split as in the random forest algorithm:

```
Load dataset
Meas=measurement data (change to Raman spectral reading)
Species=classes (change to diagnostic classes)

%Build forrest
Mdl = TreeBagger(50,meas,species,
'oobpred','On','Method','classification')

My Example
Mdl =
TreeBagger(50,bsp',class,'oobpred','On','Method','classificati
on')

%View tree
view(Mdl.Tree,'Mode','graph')

%Plot out of bag error over the number of classification trees
figure;
oobErrorBaggedEnsemble = oobError(Mdl);
plot(oobErrorBaggedEnsemble)
xlabel 'Number of grown trees';
ylabel 'Out-of-bag classification error';
```

Rattle can was also used for ensemble methods by choosing the relevant options after data has been loaded and the target variable has been identified.

4.6.6. PLS-DA

The PLS-LDA analytical methods I implemented used MATLAB and have been previously described in detail (McGregor et al., 2017) (Zhao et al., 2016). Briefly, the PLS-LDA-LOOCV procedure consisted of training and testing steps. The training process included three procedures: (1) the training spectra were standardized by subtracting the training spectra mean and dividing by the training spectra standard deviation, (2) through the NIPALS Algorithm (Hill and Lewicki, 2005), the weight factors, loadings, regression coefficient and factor scores of the training spectra were calculated, and (3) a linear discrimination model was developed to predict the classification of new cases. The testing process also included three procedures: (1) the test spectrum was standardised by subtracting the training set mean and by dividing the training set standard deviation, (2) the factor scores of the test spectrum were calculated, and (3) a posterior probability was calculated based on the linear model developed in the training step. The testing process was repeated n times such that each spectra was treated as the test spectra once. The ROC curve was calculated from the posterior probabilities derived above and represents the diagnostic performance. With good discrimination between 2 groups, the ROC curve moves toward the left and top boundaries of the graph, whereas poor discrimination yields a curve that approaches the diagonal line function. The AUCs were calculated using the trapezoidal rule (Hanley and McNeil, 1982).

4.6.7. Classifier Performance Evaluation

Evaluating the models built with experimental data is required to assess model accuracy as a basis for classifying the tissue being operated on. Common problems confronted at this stage are:

- Overfitting the data
- Balancing classes
- How best to compare models

Overfitting can be defined as when the model fits both the underlying relationship between the variables in a system and also the noise unique to each observed sample. It has implications for the generalisability of models being generated as overfitted models only predict with accuracy from the data they are trained on. One method to compensate for overfitting is to choose the correct cross validation method.

Leave one out cross validation (LOOCV) uses most of the data to create the best model, with a risk of overfitting as the data may be biased. It is limited to one observation at a time and is less generalisable when considering a new independent dataset. Leave one patient out cross validation (LOPOCV) mitigates this by leaving out a whole patient's worth of data and using that for testing, avoiding patient specific overfitting or clustering effects. The drawback of this method is that a lot of data can be left out resulting in an inadequate model to train from. K-fold cross validation is a

compromise between the 2 methods above. K fold cross validation averages the accuracy of the training model in a test group rather than a single observation. The test/training data split is a compromise between LOOCV and LOPOCV.

The results from cross validation are presented in a confusion matrix. A simple 2x2 matrix consisting of the results from an index test against the reference test, where the test outcome is binary. This is also called an error matrix. An example from one of my models is shown in Figure 34.

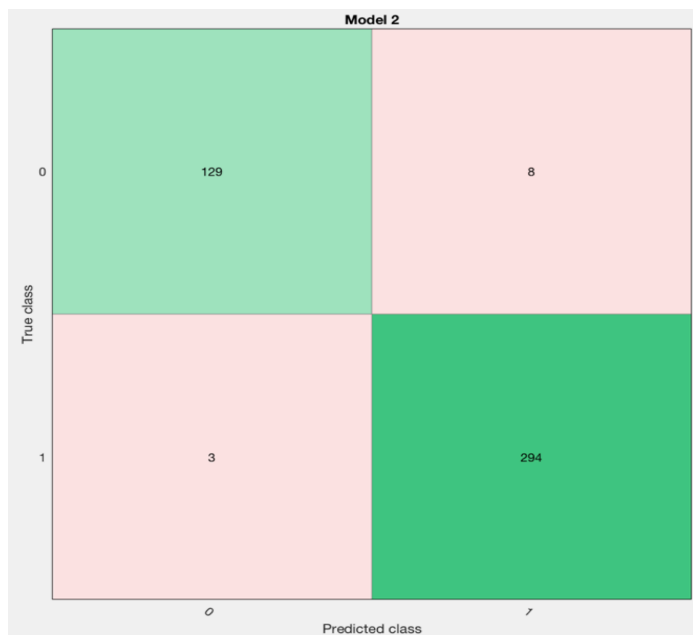


Figure 34. A 2x2 confusion matrix of a binary classifier.

Green represents correct classification, pink represents erroneous classification.

The results of cross validation can also be represented as a Receiver-Operating Characteristic (ROC) curve. A ROC curve is a graphical depiction of sensitivity (True Positive Rate) versus 1-specificity (False Positive Rate). This shows the performance of the binary classifier in the model as the threshold is varied. The Area Under the Curve (AUC) serves as the measure of accuracy of the model being tested. Figure 35 is an example of a tumour vs non-tumour model that was tested, showing an excellent AUC of 0.97. A test with perfect accuracy would return an AUC of 1.0, and a test which did not perform well would return 0.5 (a diagonal line across the graph).

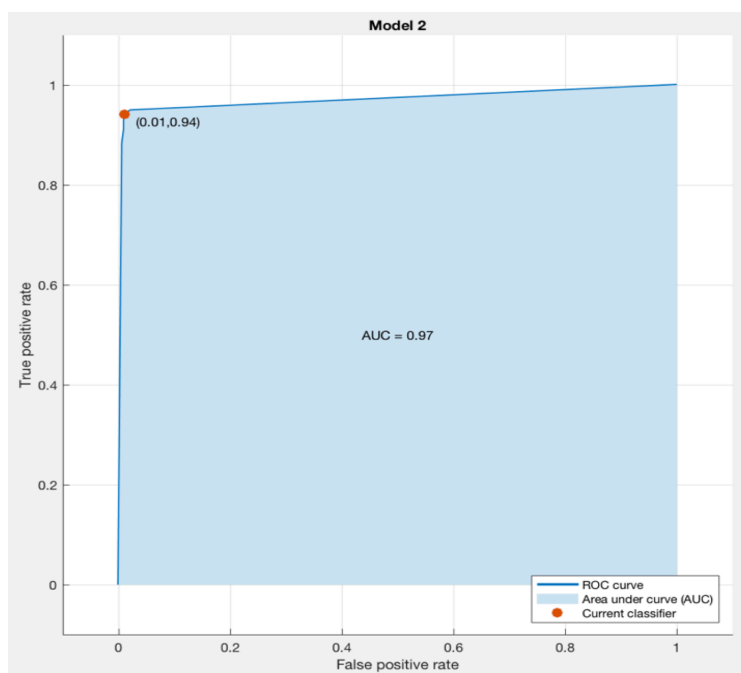


Figure 35. An example of a ROC curve calculated for tumour vs non-tumour classification using Raman data.

Area under the curve is represented in light blue, the ideal classifier is represented as a red dot. The ROC curve is in dark blue.

4.6.8. Feature Extraction/ Peak Analysis

Feature extraction is commonly performed after a model has been selected as having adequate performance. I used vector or PCA loadings to define key data features driving correct classification and then ranked and graphically represented these features. In the case of Raman spectroscopy these wavenumbers can be compared to published tissue-specific data to suggest the identity key molecular bonds and regions of the Raman spectrum which are responsible for driving the classifier.

4.6.9. Summary

Analysis of Raman data requires powerful data manipulation and analysis techniques and utilises many branches of machine learning as detailed in the sections of this chapter. I have defined data pre-treatment, model building and testing and model selection as key elements of interpreting Raman spectra in order to classify disease. A brain-specific data analysis platform using code in Matlab and 'R' has been shown here which aims to extract key molecular profiles for various disease states.

In the following chapters the *in vivo* results of using machine learning from Raman spectroscopy will be presented.

4.7. Raman Spectroscopy Results

4.7.1. Introduction

The following chapters present the main results I obtained using the *in vivo* Raman system described in section 3.4. The data will be presented first as a demographic summary of patients and samples from which the data has been obtained followed by a more detailed account of the first brain *in vivo* spectra acquired during brain tumour surgery in Europe with key classification models being presented together with model performance measures. This will provide an overview of the utility of the molecular information I obtained using *in vivo* Raman spectroscopy in brain tumour surgery.

4.7.2. Demographics

4.7.2.1. Introduction

The Raman spectroscopy study successfully recruited a total of 33 patients. Of these, 4 patients were excluded from the analysis. The reasons for exclusion included 1 patient who was too unwell for surgery and research equipment failure in 3 patients (in 2 cases the probe did not pass fibre optic testing and in 1 case the Windows operating system crashed requiring reinstallation). This left 29 patients for which complete data sets were collected.

Metastases are the commonest tumour of the brain. The commonest intrinsic brain tumour arising from the brain substance itself is the highly aggressive Glioblastoma Multiforme (GBM), followed by astrocytomas and oligodendrogliomas in frequency. The data presented here follows these general trends observed in larger populations.

4.7.2.2. Demographic analysis

The 29 patients recruited represent a cross section of brain tumour patients seen by the neurosurgery service. The majority were high grade gliomas, with some low grade gliomas, metastases and rarer brain tumours included. The patients were recruited in a consecutive series depending on equipment availability. A demographic summary table is shown in Table 7.

P Number	Gender	Age	iKnife	Histology Result	WHO Grade
1	M	34	1	Astrocytoma	2
2	M	58	0	GBM	4
3	M	22	1	GBM	4
4	M	22	1	Astrocytoma	2
5	M	30	1	Astrocytoma	2
6	M	42	1	GBM	4
7	M	34	0	GBM	4
8	M	73	0	GBM	4
9	F	73	0	GBM	4
10	M	49	0	Astrocytoma	2
11	M	42	1	Astrocytoma	3
12	M	63	1	Haemangioblastoma	N/A
13	M	37	0	Astrocytoma	2
14	F	57	1	Choriocarcinoma	N/A
15	F	48	1	GBM	4
16	M	79	1	GBM	4
17	F	56	0	Lung met	N/A
18	F	37	1	Meningioma	1
19	F	52	1	Astrocytoma	3
20	F	48	0	Meningioma	1
21	M	38	0	Oligodendroglioma	3
22	M	48	1	GBM	4
23	F	38	1	GBM	4
24	M	36	0	Oligodendroglioma	2
25	M	70	1	GBM	4
26	F	33	1	Oligodendroglioma	2
27	M	53	1	GBM	4
28	M	52	1	GBM	4
29	M	38	1	Radionecrosis	N/A

Table 7 Summary table of all patients recruited to the Raman spectroscopy study.

29 patients were successfully recruited. Data collected included gender, age, histology result and WHO grade of tumour as well as whether patients were also recruited as part of the iKnife study.

This data is represented graphically in Figure 36 below.

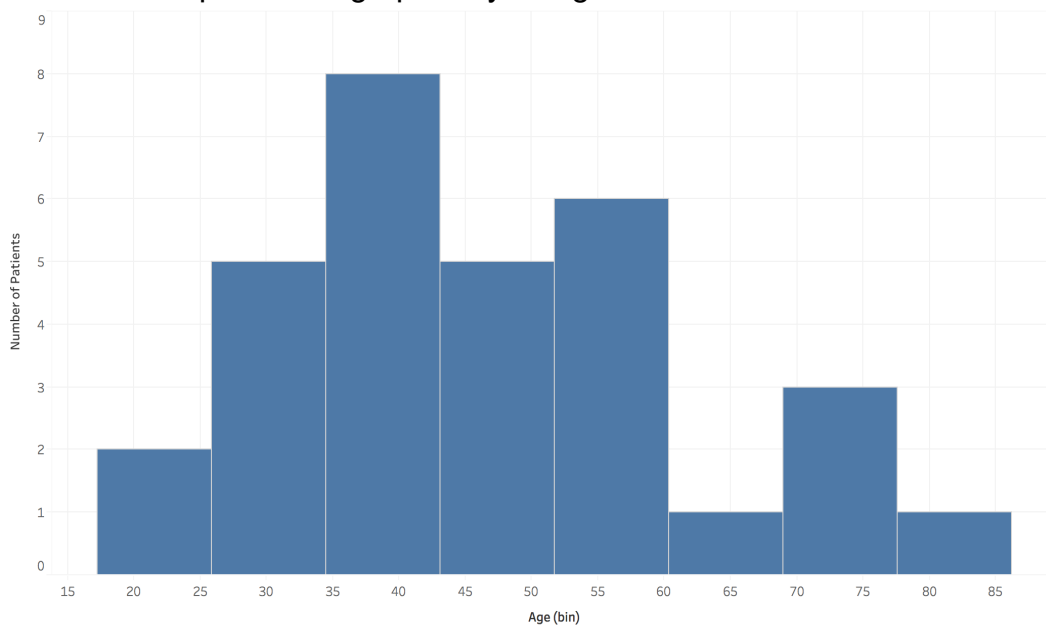


Figure 36. Histogram representing age distribution of recruited patients.

The age distribution is slightly skewed towards the 35-45 age range with a long tail in the distribution towards more elderly patients. This reflects our patient practice. Males were recruited more frequently than females as shown in Figure 37.

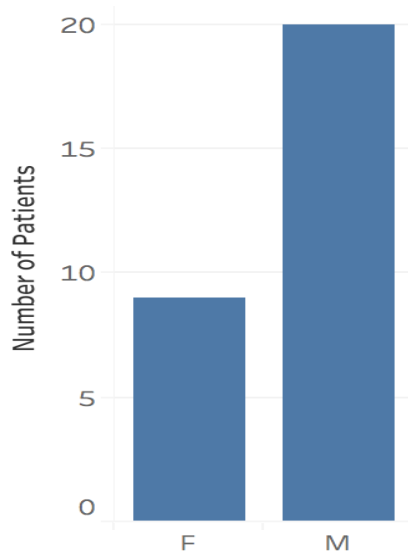


Figure 37. Number of patients recruited for each gender.

As expected from previously existing public health data, GBM was the commonest tumour found, followed by astrocytoma and oligodendroglioma as shown in Figure 38. Metastases, meningiomas and tissues with radio necrosis represent non-glioma tissue. A total of 6 patients with brain tumours that were non-glioma were recruited which consisted of 2 patients meningioma, 1 choriocarcinoma, 1 haemoangioblastoma, 1 lung metastasis and 1 patient with radionecrosis. This represents a good cross section of intrinsic and extrinsic brain tumours which are operated on.

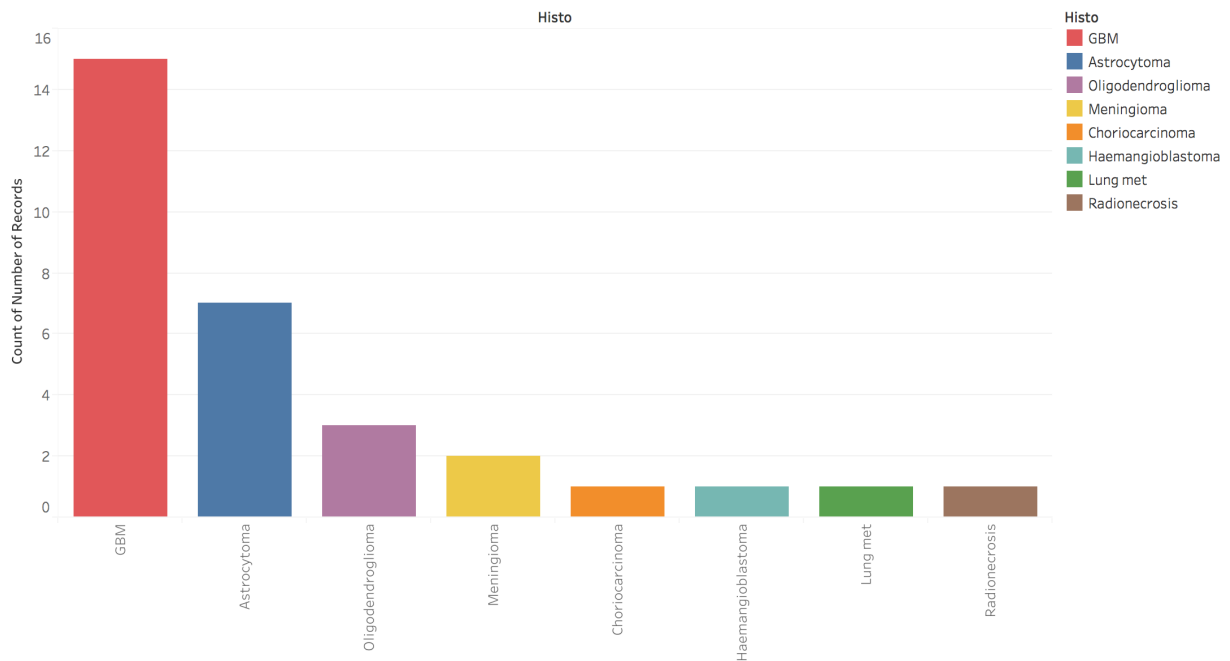


Figure 38. Distribution of histology findings amongst recruited patients. *Histological diagnoses arranged in order of frequency. GBM, astrocytoma and oligodendroglioma form the glioma group of tumours. Meningioma, choriocarcinoma haemoangioblastoma and lung metastasis and radionecrosis form the non-glioma lesions.*

Patients recruited into this study had tumours spanning all 4 WHO grades with WHO grade 4 being the most common, as shown in Figure 39.

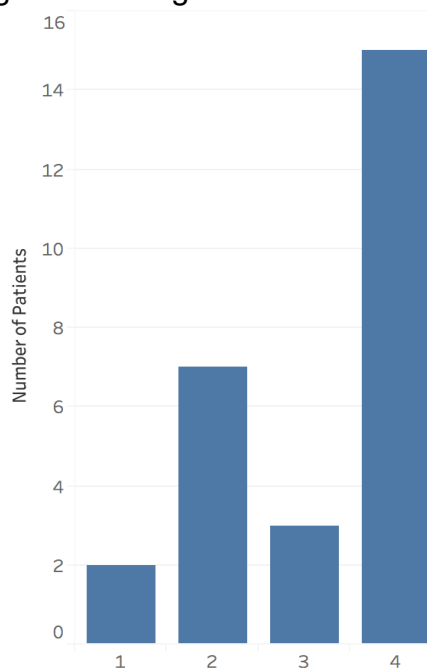


Figure 39. Distribution of WHO grades (2016 classification) amongst recruited patients. *WHO clinicopathological grades arranged in numerical order. The highest number of cases were WHO grade 4, a very aggressive, incurable form of disease.*

4.7.2.3. Raman spectroscopy summary: spectral data

The total number of spectra I collected, combining *in vivo* and *ex vivo* spectra, was 2,437. This may be one of the largest brain Raman spectral databases collected. The majority were spectra I obtained on *ex vivo* tissue which was submitted for histopathological analysis (1266) with 1171 spectra obtained *in vivo*. *in vivo* tissue spectra matched with biopsies totalled 300, leaving 971 spectra without biopsies. Figure 40 summarises the spectral demographics.

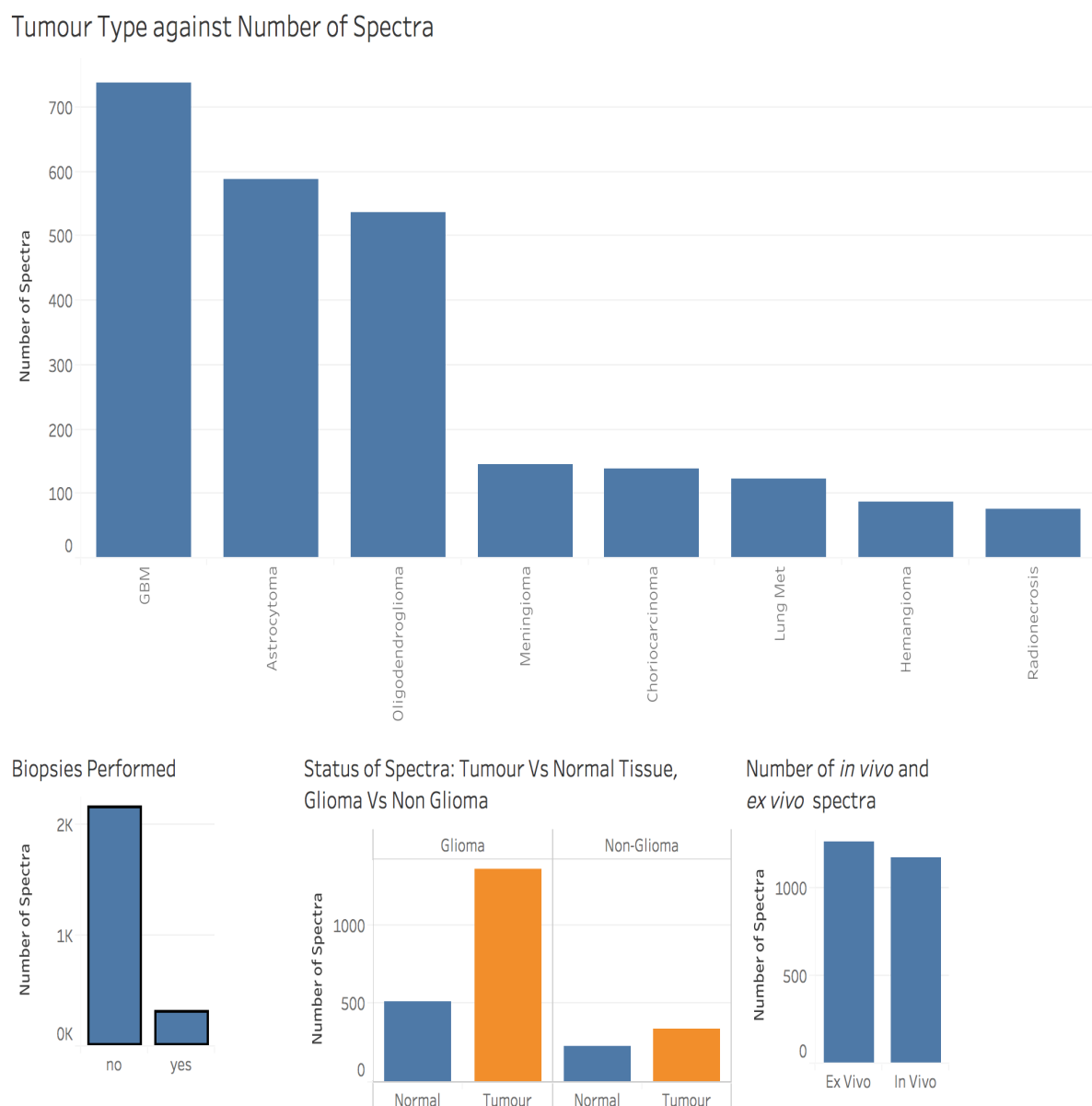


Figure 40. Demographics of the spectra collected in the study.

The number of spectra for each tumour type, in vivo vs ex vivo spectra, normal vs tumour, glioma vs non-glioma and number of in vivo biopsies performed. Note: all ex vivo spectra were obtained from histopathologically verified biopsy tissue.

The *in vivo* spectra spanned all tumour grades and tumour types and were obtained from 25 patients. Table 8 summarises these spectra.

WHO Grade	Tumor Type	# Of Patients	# Of Normal Spectra	# Of Tumor Spectra	Total Spectra
G2	G2 Astro	5	130	105	235
	G2 Oligo	2	63	72	135
	G2 Oligo/GBM	1	35	12	47
	G2 Type Totals	8	228	189	417
G3	G3	1	32	23	55
	G3 Oligo	1	0	9	9
	G3 Oligo (GBM)	1	37	35	72
	G3 Oligoastro	1	35	0	35
	G3 Type Totals	4	104	67	171
G4	GBM	8	104	261	365
G2,G3 and G4 Type Totals		20	436	517	953
Other	Hemangioma 4C	1	23	0	23
	Lung Met	1	32	0	32
	Meningioma	1	38	0	38
	Radionecrosis	1	48	0	48
	Choriocarcinoma	1	38	39	77
	Grade Totals	5	179	39	218
All Grades	All Types	25	615	556	1171

Table 8. Distribution of numbers of *in vivo* spectra amongst different tumour grades and diagnosis types.

All tumour grades and tumour types had Raman spectra collected. Meningiomas in this study were all classified as WHO Grade 1. A total of 1171 in vivo spectra were obtained.

Of the 1171 spectra I obtained during surgery, a roughly equal split was obtained between tumour spectra and normal spectra. In the vast majority of cases, I obtained normal spectra by measuring Raman spectra as far away from tumour as possible in areas which were macroscopically and radiologically normal. Performing a biopsy in these areas was contraindicated due to the possibility of this being functional normal brain and would be a significant departure from accepted practice. In some cases of uncertainty, biopsies were performed which later showed normal tissue.

4.7.2.4. Summary

I have collected one of the largest Raman datasets for brain tumours which has been summarised in this chapter. This can be analysed in many ways. In the following sections, the results I obtained are organised with respect to the following clinical questions, hypotheses, and the relevant data experiment to explore if a successful model can be built to classify the data into clinically useful groups:

1. Can *in vivo* and *ex vivo* spectra be used interchangeably for model building?
Aim: to increase the data available for model building by combining *in vivo* and *ex vivo* data.
2. Can Raman spectroscopy be used to distinguish between normal and tumour tissue?
Aim: to develop a clinically useful classifier for *in vivo* surgical decision making when resecting tumour tissue.
3. Can Raman spectroscopy be used to distinguish tumour type?
Aim: to inform the surgeon during surgery as early as possible about the diagnosis.
4. Can Raman spectroscopy be used to distinguish intrinsic brain tumour grade?
Aim: to inform the surgeon as to the prognosis of the patient.
5. Can Raman spectroscopy be used to detect tumour density?
Aim: to develop a cut off tumour density for resection of high grade glioma.
6. Can Raman spectroscopy be used to detect tumour margins?
Aim: to explore the possibility of Raman spectroscopy as a margin control tool.
7. Can Raman spectroscopy be used to detect the genetic status of tissue?
Aim: to explore the ability of Raman spectroscopy to gather clinically useful molecular information about tumour genotype.

4.7.3. *in vivo* Raman Spectroscopy of the Brain

4.7.3.1. Introduction

There only one published example of *in vivo* Raman spectra of human brain tissue (Jermyn, 2015). This chapter aims to present the raw spectra I obtained from the Raman pilot study and the first few spectra obtained after processing, showing the effect that processing has on transforming raw data into more usable Raman spectra. Some of the observed characteristics of the Raman spectra will be discussed.

4.7.3.2. Spectral Processing and Features

Figure 41 shows an example of a calibrated and smoothed raw emission spectrum I obtained *in vivo* from a brain tissue site of a patient enrolled in the study. Above 400 cm^{-1} the figure shows the calibrated and smoothed raw emission as a superposition of Raman emission peaks and a slowly varying auto-fluorescence. The iterative polynomial fits I used in this range to extract these two components are also shown. Below 400 cm^{-1} the calibrated raw emission intensity drops rapidly. This is partly due to the action of the system's long pass transmission filters, but also because there

was likely to be some small background component that could not be fully removed which led to a systematically smaller calibration correction than was required. The process of obtaining Raman spectra depicted graphically in Figure 41 was repeated for all 1171 *in vivo* spectral measurements.

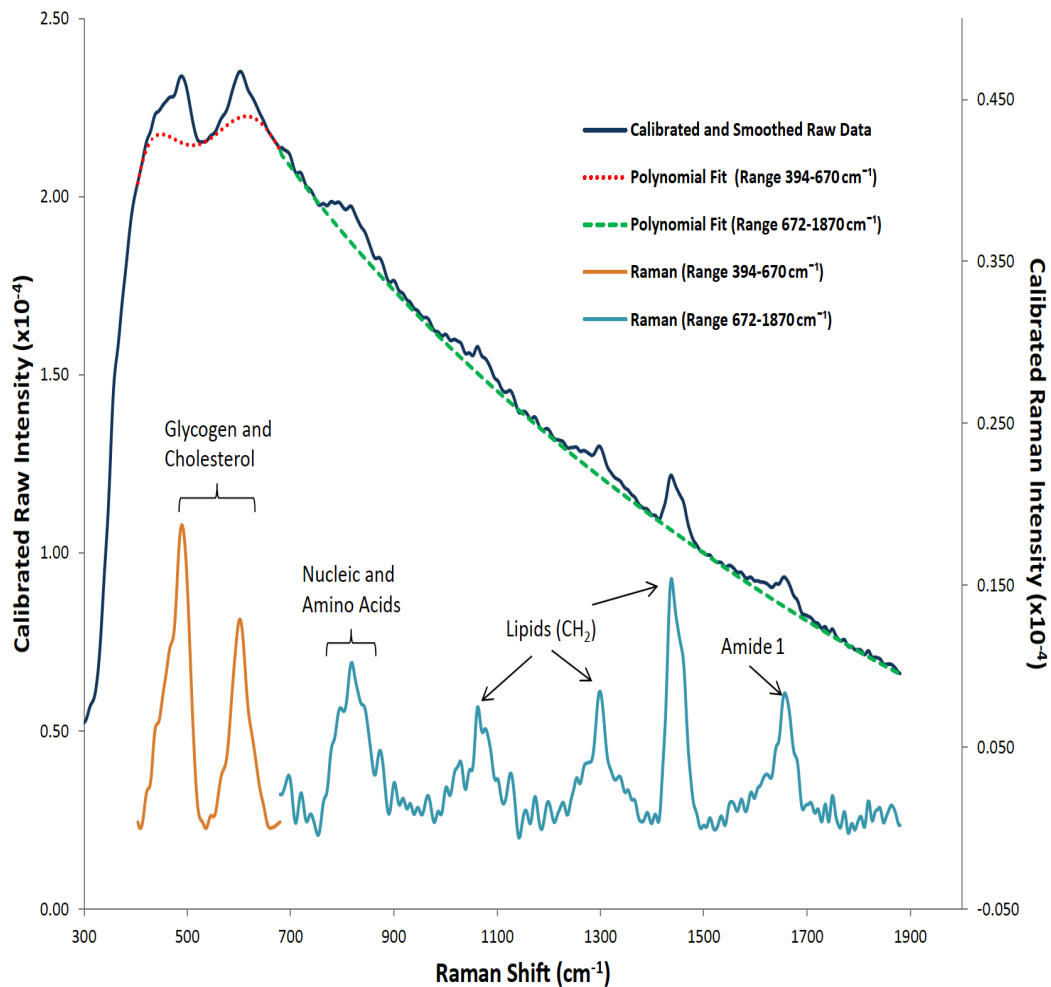


Figure 41. Example of a two range (394 – 670 cm⁻¹ and 672 – 1870 cm⁻¹) iterative polynomial fit to a calibrated and smoothed tissue emission spectrum.

The raw data was acquired by an in vivo brain tissue measurement of a visually normal site in a patient recruited in this study. Post-operatively the raw spectrum was calibrated and smoothed as described in the methods section, and the pathology of a biopsy later confirmed the patient had a G2 Oligo tumour.

Example *in vivo* Raman spectra produced after background subtraction, calibration and smoothing of the data are shown in Figure 42 below. These spectra show differences between the grades of tumour based on the WHO grading scale and illustrate a number of features.

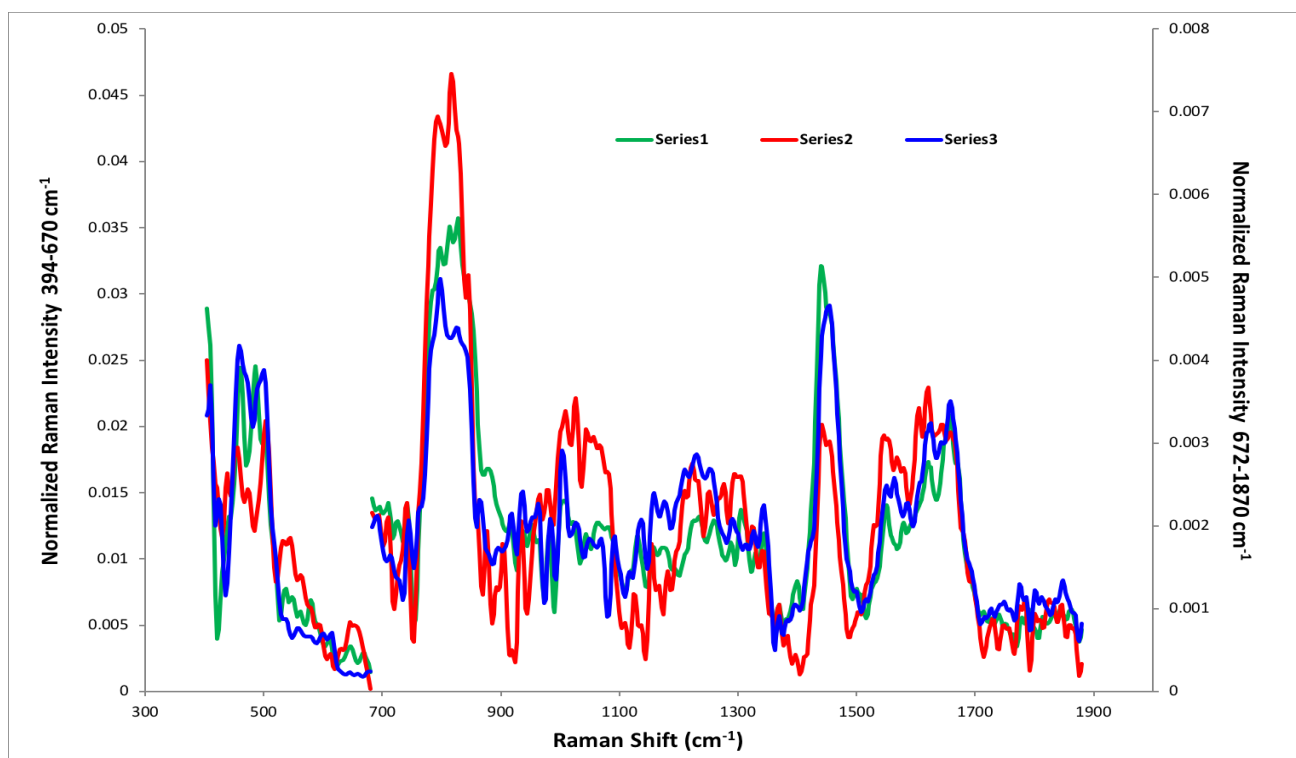


Figure 42. Raman spectrum of in vivo brain tissue of different WHO grades.

Example Raman spectra from range 394-1870 wave numbers. A break at 700nm is seen due to optical filters used in the system. Series 1=Grade 2, Series 2-Grade 3 and Series 3-Grade 4 tumours on the WHO grading scale.

Figure 42 shows that data in all areas of the Raman range were sampled from 394-1870. There is significant information in the DNA range, the lower end of the spectral range, and in the lipid and protein regions. In particular most brain spectra show the following clusters of peaks:

Broad peaks at 800 and 1450

Multiple peaks at 1500-1700

Significant differences between tumour grades throughout the whole spectrum (DNA/protein/lipid regions)

It appears that the whole spectral range may carry diagnostic information.

4.7.3.3. Summary

Raman spectra from tumour tissue were found to be unique with differences distributed across the whole spectral range. This data will now be analysed to investigate whether significant differences exist between *in vivo* and *ex vivo* spectra, tumour and non-tumour spectra, different tumour types and grades, different tumour densities and the genetic status of the tumour tissue. Predictive models will be built from training data sets and these will be applied to testing data sets to measure their accuracy.

4.7.4. *in vivo* vs *ex vivo*

4.7.4.1. Introduction

The main questions I sought to answer from data experiments which follow were:

1. How similar are the *in vivo* and *ex vivo* datasets? This has implications on whether data from *ex vivo* tissue can be used to create or augment an *in vivo* database. This then led to 2 further questions:

2. Can *ex vivo* data be used to predict *in vivo* data with accuracy? Important implications for improving system accuracy in different environments and make the system useful for a broader range of patients, important for multi centre trials.

3. Does a combined analysis of *ex vivo* and *in vivo* data increase accuracy? This combined analysis was performed for all the main domains of analysis: Tumour vs Normal, Tumour Type (Glioma, Non-Glioma, Metastasis), Tumour Grade (WHO grading) and Tumour Cell Density (Normal brain, infiltrated brain, solid tumour tissue).

The hypothesis is that the molecular nature of non-fixed *ex vivo* tissue will be similar to *in vivo* tissue as tissue fixation methods will not have been used. There will therefore be a negligible difference between the spectra.

4.7.4.2. Multivariate analysis

In total the following number of spectra were collected for analysis:

in vivo 1171 spectra

ex vivo 1266 spectra

The spectral demographics are summarised in Figure 43.

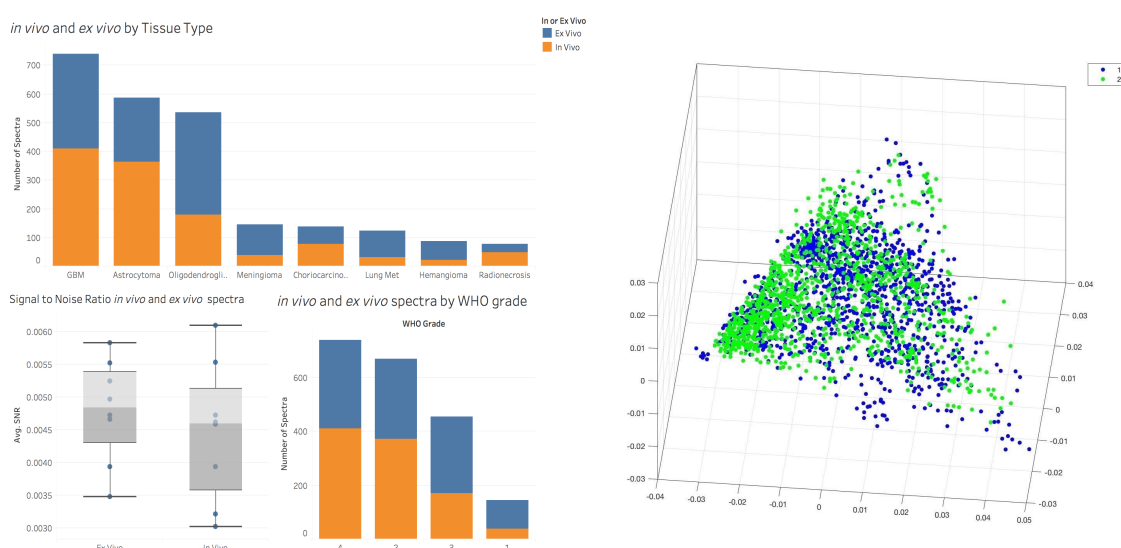


Figure 43. Characteristics of spectra collected from *in vivo* and *ex vivo* sources.

The distribution of *in vivo* and *ex vivo* spectra amongst tumour types (top), tumour grades (bottom) and respective average Signal to Noise Ratios (SNR). 3-D representation of PCA of *in vivo* vs *ex vivo* Raman spectra (right). The first 3 principle components are presented here. 1=*in vivo* spectra 2=*ex vivo* spectra. There is no convincing separation between *in vivo* and *ex vivo* data.

The spectral demographics show *in vivo* and *ex vivo* spectra were obtained in all tumour types and grades with no significant difference in signal to noise ratio. This shows that spectra were obtained from a wide variety of tumour types and grades. There is a slightly higher variance of SNR when spectra were recorded *in vivo* which may be due to the more variable sampling conditions when recording spectra in situ. The SNR from *in vivo* and *ex vivo* data is acceptable in terms of absolute values. The spectra were analysed first using principle component analysis as an unsupervised method as shown in figure 43. The process for performing PCA has already been described in section 3.6.3.

I then performed supervised data analysis using LDA as previously described in section 3.6.3 and is shown in figure 44.

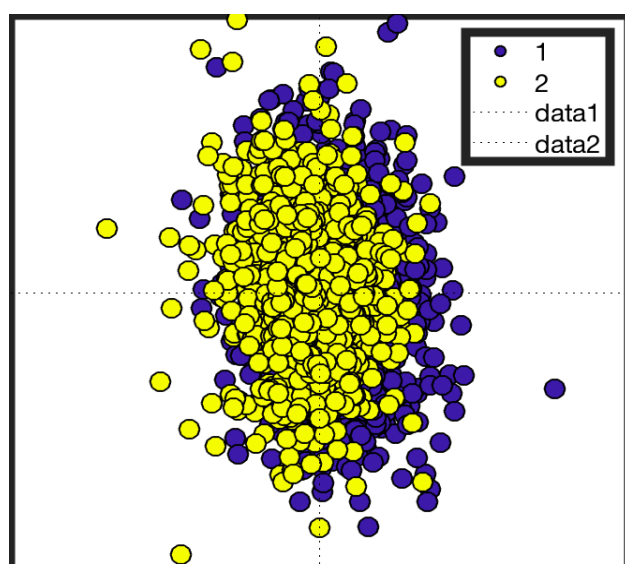


Figure 44. LDA plot of *in vivo* and *ex vivo* data.

1= *In vivo*, 2=*Ex vivo*

This LDA plot predicts shows that *in vivo* and *ex vivo* data overlap greatly, predicting poor ability to classify between the 2 groups.

My analysis of *in vivo* and *ex vivo* data using machine learning showed poor ability to build a model to discriminate between *in vivo* and *ex vivo* data, as shown in the decision tree model below. Of all machine learning tools employed, only SVM analysis produced a model with some ability to differentiate between *in vivo* and *ex vivo*, but this was not very accurate.

The error matrix for the SVM model showed poor performance:

Actual	Predicted	
	Ex Vivo	In vivo
Ex Vivo	221	38
In vivo	34	195

Error matrix for the SVM model on Rattle Invivo and Exvivo combined.csv [test] (proportions):

Actual	Predicted		
	Ex Vivo	In vivo	Error
Ex Vivo	0.45	0.08	0.15
In vivo	0.07	0.40	0.15

Overall error: 15%, Averaged class error: 15%

Rattle timestamp: 2018-04-04 12:55:25 Babar Vaqas

=====

4.7.4.3. Summary

On the basis of this analysis, combined data set was analysed for each of the main outcomes of the study, as it seems apparent that there are no discernible differences between the 2 data sets. PCA and LDA analysis did not show any clustering of the data. All the machine learning methods failed to find differences apart from a very weak performing SVM model.

4.7.5. Tumour Vs Normal

4.7.5.1. Introduction

In the previous results chapters the spectra, samples and patient characteristics of the core dataset have been established. I determined that predictive models cannot be trained to separate *in vivo* and *ex vivo* data using unsupervised or supervised machine learning methods. These data groups will therefore be combined for the following analyses to ensure the training and testing data sets are as large as possible for model building.

The most important clinical question is whether Raman spectroscopy when used *in vivo* can distinguish between tumour and non-tumour tissue in order to supplement the technologies already used summarised in section 2.3.2. The results of my analysis will be presented here.

4.7.5.2. Average spectra

A total of 649 normal spectra and 522 tumour spectra were included in this analysis. This represents all of the *in vivo* spectra collected. Figure 45 shows the averaged normal and tumour spectra. The difference spectrum is also represented with significant differences highlighted in red based on the Student T test. It can be seen that differences exist across the whole range of the Raman spectrum.

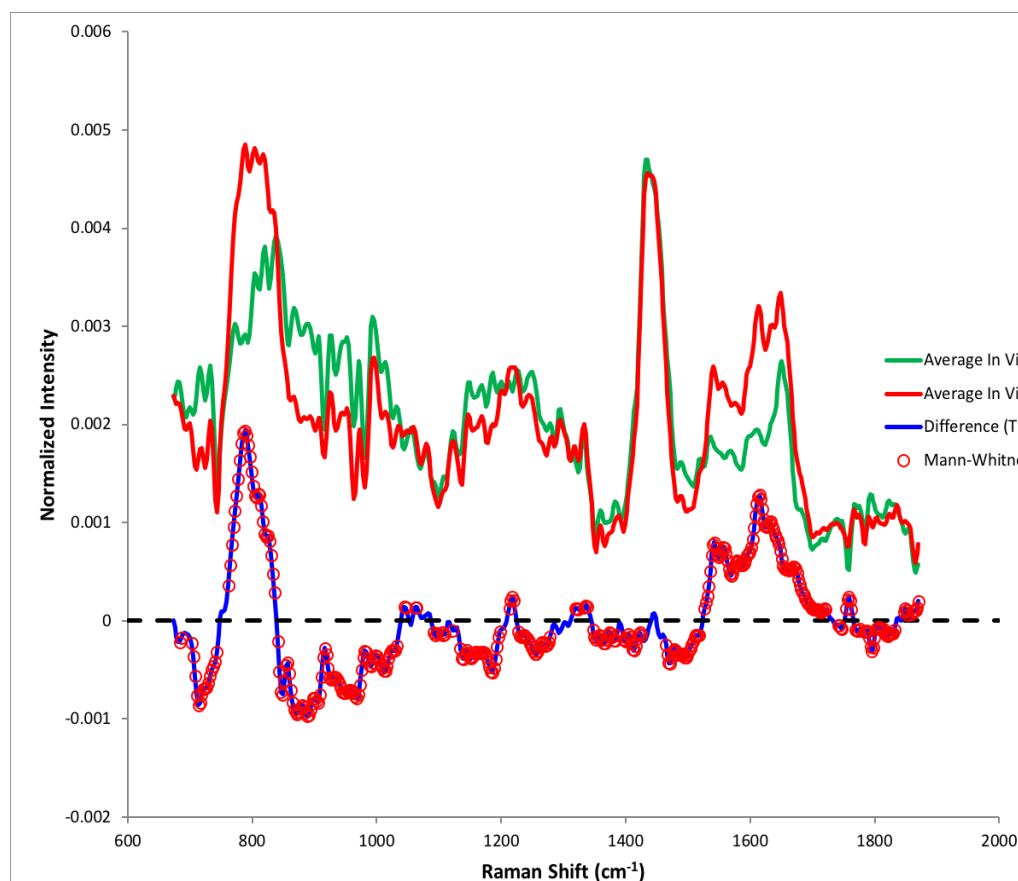


Figure 45. Average normal vs average tumour spectrum.

The difference spectrum is shown below in green. Significant differences ($P < 0.05$ Student *T* test) are shown in red open circles. Multiple areas of significant difference exist throughout the 600-1900 wavenumber range, with the normal/tumour spectra being showing major differences on visual inspection.

The most significant high intensity differences are in the 750-850 wavenumber range, 900-1000 range and the 1500-1700 range. Interestingly there is a strongly conserved high intensity peak at 1450 cm⁻¹ which is shared between tumour and normal tissue with no significant difference.

In order to potentially identify the significance of these regions in the *in vivo* brain spectrum, as a crude test, I undertook a peak comparison with known tissues and molecules as in Figure 46.

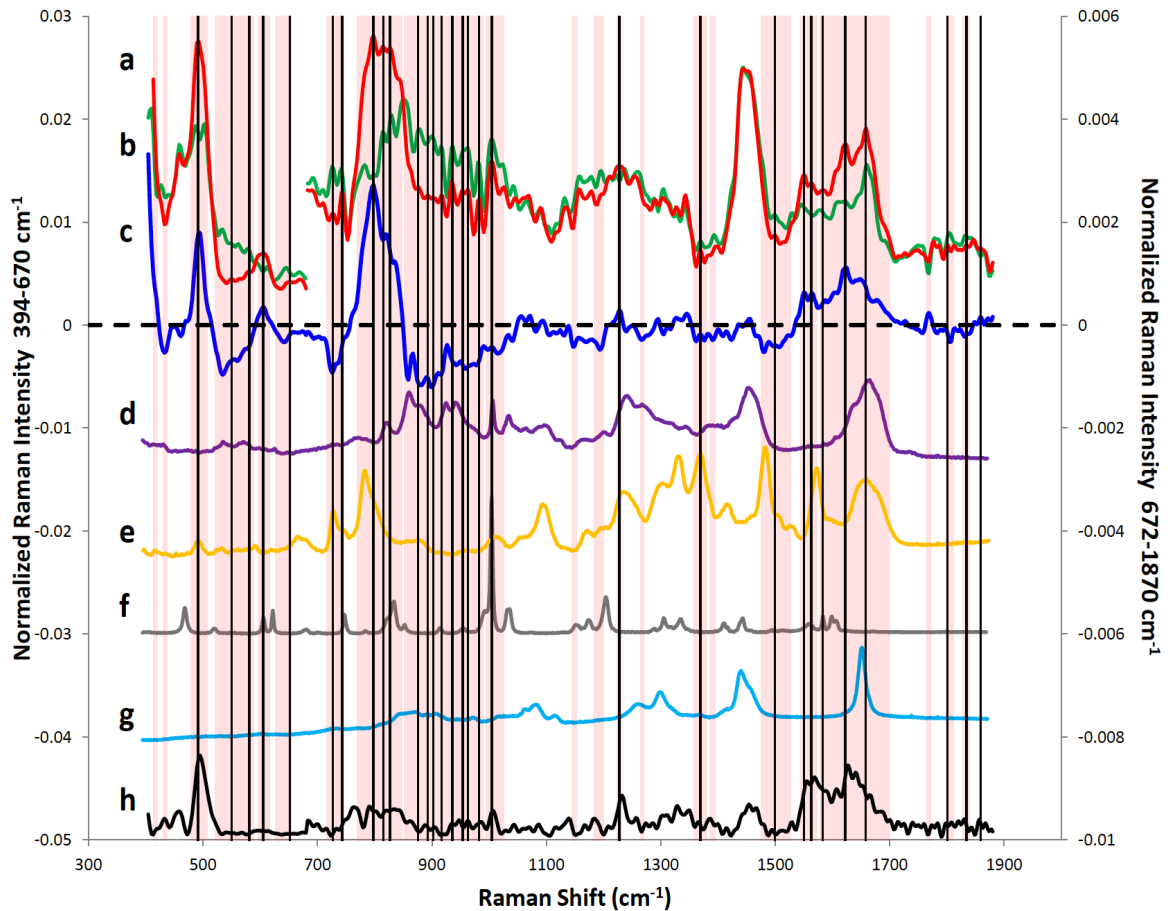


Figure 46. Comparison of normalized *in vivo* Raman spectra from brain tissue and Raman spectra from biochemical standards expected to be found in brain tissue in abundant quantities.

Tissue spectra were normalized by the area under the curve in two ranges: 394 – 670 cm^{-1} and 672 – 1870 cm^{-1} and plotted on two different intensity scales for clarity. From top to bottom the spectra are: a) average of all tumour spectra ($n = 556$), b) average of all normal spectra ($n = 615$), and c) difference between the tumour and normal averages. The biochemical standards were: d) DNA, e) RNA, f) Collagen, g) Lipid (Oleic Acid), and h) Human Blood (all standard spectra were shifted on the intensity scale for clarity). Shaded wavebands are ranges over which a non-parametric statistic (Mann-Whitney U) indicated a highly significant ($p < 10^{-6}$) difference between the tumour and normal spectra. Vertical lines within these wavebands identify the location of peaks within the tumour average spectrum.

The Raman spectra, normalized then averaged in all tumour and all normal pathology classes, are shown above. The difference spectrum between these two averages is also shown in the Figure 46 along with the Raman spectra of some biochemical standards expected to be present in significant amounts within brain tissue. From this comparison it can be inferred that the bands identified at the lower end of the spectrum (750-850) are shared with RNA, the bands in the mid region (900-1000) are shared with DNA and the bands in the higher region are shared with lipid, RNA and DNA. Interestingly, the highly conserved peak near 1450 is likely to be a lipid peak.

4.7.5.3. Multivariate analysis

I performed exploratory unsupervised multivariate analysis in order to investigate patterns in the data to guide subsequent model building and the to be groups included in the supervised analysis.

Simple PCA including all 1171 spectra obtained *in vivo* did not show any convincing separation between normal and tumour spectra as shown in Figure 47. This could be because 'normal' samples identified using MRI for GBM are actually tumour. To test this hypothesis further analyses were undertaken using biopsy verified histology only and analysis was undertaken within WHO 2-3 disease groups. Grade 2 and to a lesser extent grade 3 disease is more focal with less field change in adjacent normal tissue.

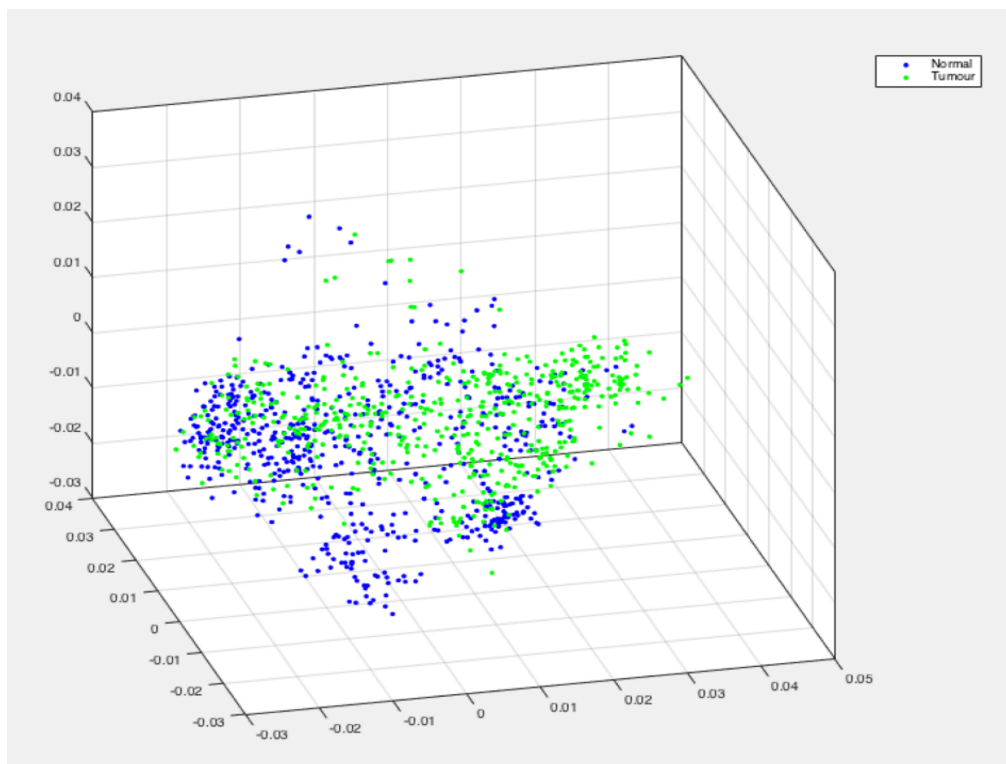


Figure 47. PCA of all 1171 spectra from brain tissue obtained during surgery. Blue dots are normal, green dots are tumour spectra. There is no clear clustering seen between the 2 groups.

Performing PCA analysis on spectra with biopsy data only still shows poor separation because of the low numbers of biopsy proven normal tissue as shown in Figure 48. Using this as a training data set is likely to result in over classification of tumour spectra.

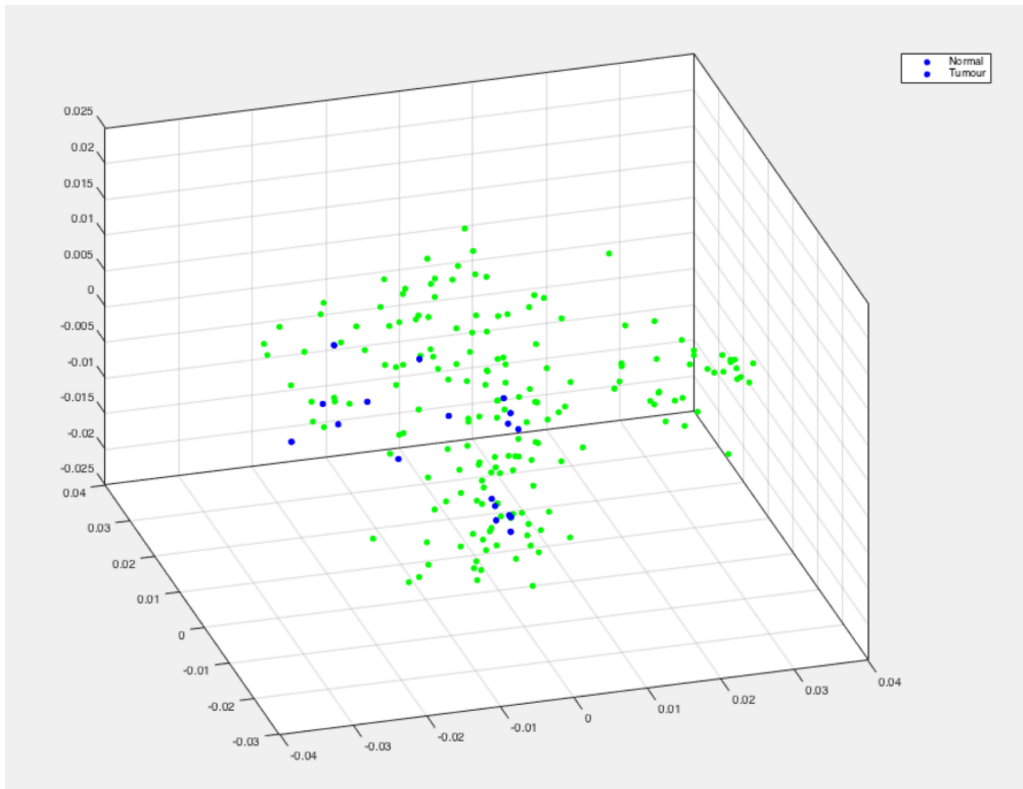


Figure 48. PCA of Raman spectra with biopsy information only.

Lack of clear separation on PCA between tumour (green) and normal (blue) with few normal data points making comparison difficult.

PCA was then performed on all in vivo spectra but leaving out GBM data. This produced a better separation of tumour and normal spectra as shown in Figure 49.

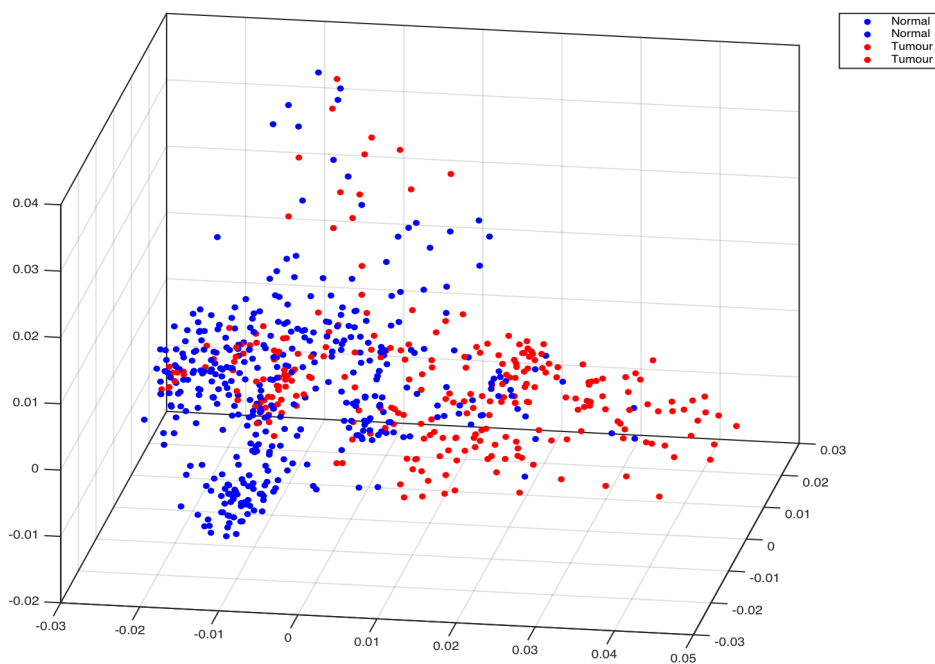


Figure 49. PCA of tumour vs normal *in vivo* spectra, excluding GBM.

Normal spectra are blue dots, tumour spectra are red dots. Some separation can be seen between tumour and normal spectra indicating a fundamental difference in the data based on unsupervised analysis alone.

Due to this difference observed between WHO grades in terms of separation of tumour and normal spectra analyses were performed with respect to the different WHO groups for subsequent model building and supervised analysis.

Supervised multivariate statistical analyses of the 1171 Raman spectra showed a clear separation of the all tumour and all normal pathology classes with a trend towards greater separation for later stage cancers (Figures 50, 51 and the inset of Figure 52). A receiver operator characteristic (ROC) was generated from the posterior probability distribution and is shown in Figure 52. The area under the ROC curve as a percentage of the total was 92% (95% confidence interval (CI) of 90-93%), and for a nominal sensitivity of 85% (95% CI of 82-88%), a specificity of 82% (95% CI of 79-85%) was obtained. Out of the three analyses methods used, PLS with selected wavebands produced marginally better results than the PLS without waveband selection, or the boosted trees method (see Table 9).

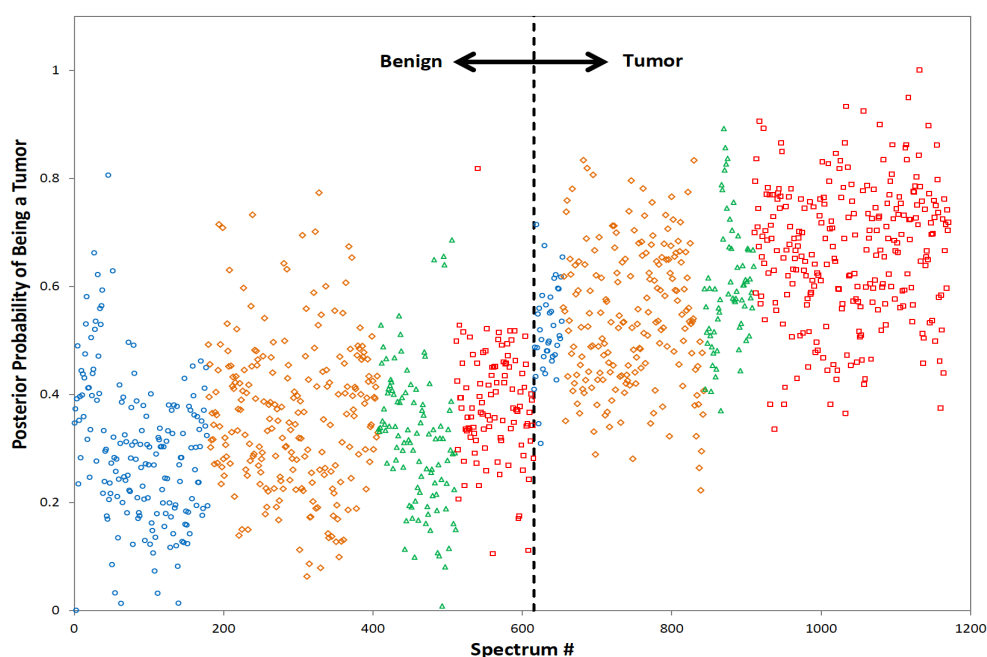


Figure 50. Posterior probability plots obtained from the analyses of 1171 Raman spectra using PLS with waveband selection.

Each symbol represents the result of one spectrum and the different symbols represent data obtained from patients with the following WHO grade tumours: Blue circles= G2, orange diamonds= G3 and red squares= G4 (GBM). Green triangles= spectra taken from patients with other brain tumour types.

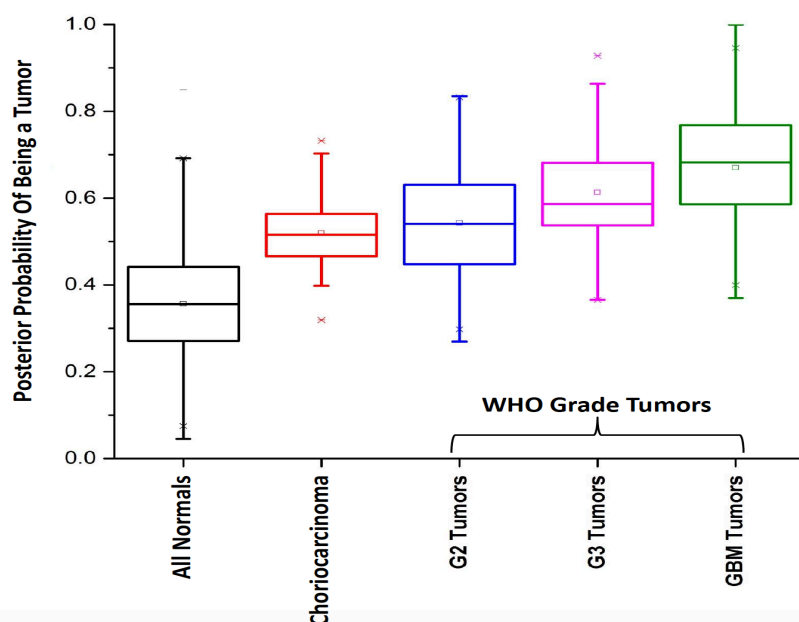


Figure 51. Boxplot representations of the posterior probability distributions as determined by PLS with waveband selection analyses of 1171 Raman spectra. The lower bound on each box shows the 25th percentile, where the upper bound shows the 75th percentile, meaning 25% and 75% of the data points are below these two bounds respectively. The line between these two bounds is the median. The whiskers on the plot show the 5th and 95th percentile, meaning 5% and 95% of the data are below these whiskers respectively. Data found outside the 5–95th percentile whiskers are outliers, as shown by separate data points.

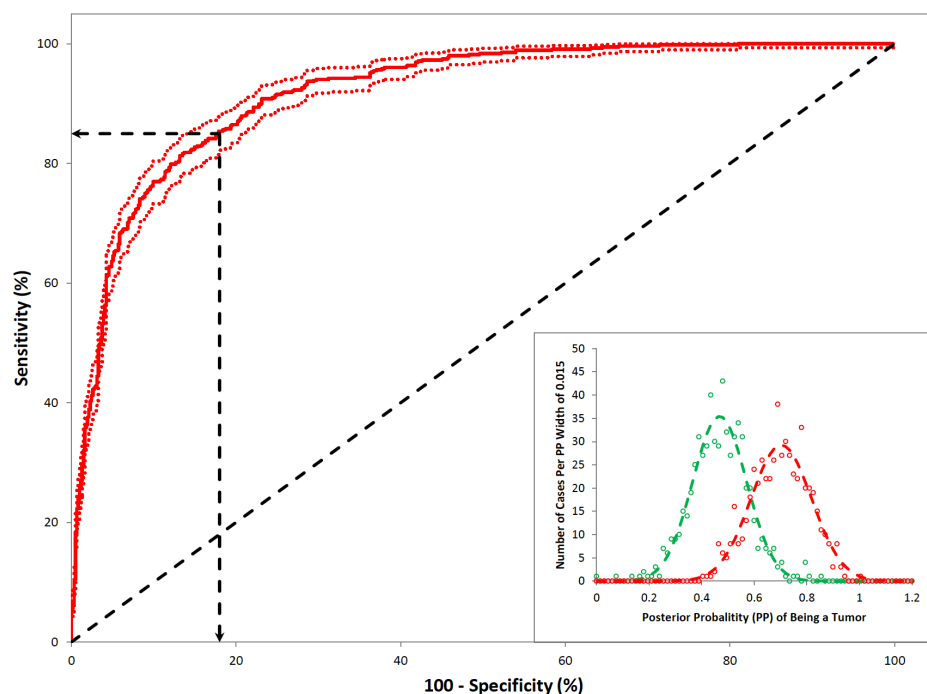


Figure 52. The ROC curve (continuous curve) and 95% CIs (dashed curves) derived from the posterior probabilities determined by PLS with waveband selection analyses of 1171 Raman spectra. The diagonal dashed line is the random guess line, and the horizontal and vertical lines intercept the ROC curve at a sensitivity of 85% (95% CI of 82–88%), a

specificity of 82% (95% CI of 79-85%). The area under the curve as a percentage of the total is 92% (95% CI of 90-93%). Inset: frequency distribution in posterior probability values fitted with Gaussian models (Green circle= normal and red circle= tumour) showing good separation of normal and tumour spectra.

Analyses Results		Test Data	G2, G3 and G4 Cases Only		All Cases	
Training Data	Method	AUC (%)	Spec. (%)	Sens. (%)	Spec. (%)	Sens. (%)
All Raman Spectra	Boosted Trees	89 (88 to 91)	67 (62 to 71)	87 (84 to 90)	72 (68 to 76)	85 (82 to 88)
	Partial Least Squares Full	89 (87 to 90)	71 (67 to 75)	85 (82 to 88)	74 (70 to 77)	85 (82 to 88)
	Partial Least Squares Select	92 (90 to 93)	80 (75 to 83)	85 (82 to 88)	82 (79 to 85)	85 (82 to 88)
	Boosted Trees	94 (93 to 96)	83 (79 to 86)	87 (84 to 90)	86 (83 to 89)	85 (82 to 88)
All Raman + Fluorescence Spectra	Partial Least Squares Full	92 (91 to 94)	83 (79 to 87)	87 (83 to 89)	83 (80 to 86)	85 (82 to 88)
	Partial Least Squares Select	93 (92 to 94)	86 (82 to 89)	86 (82 to 89)	87 (84 to 90)	85 (82 to 88)
	Boosted Trees	94 (93 to 96)	83 (79 to 86)	87 (84 to 90)	86 (83 to 89)	85 (82 to 88)

Table 9: Classification of spectra acquired from patients with different tumour subclasses based on different multivariate methods.

Percentages in parentheses are the 95% confidence intervals. Classification was based on training and test spectra drawn from the full data set of 1171 normalized spectra. Two classifications were performed a) using Raman spectra only and b) using combined Raman and Fluorescence spectra.

Table 9 also shows that fluorescence signal increases the accuracy of all models (boosted trees, PLS with and without waveband selection). It is likely that there is diagnostic information in the autofluorescence measured from brain tissue with increased specificity and negligible effect on sensitivity. I investigated this fluorescence effect further.

The auto-fluorescence spectra from the 1171 measurements were averaged by all tumour and all normal pathology classes (Figure 53 below).

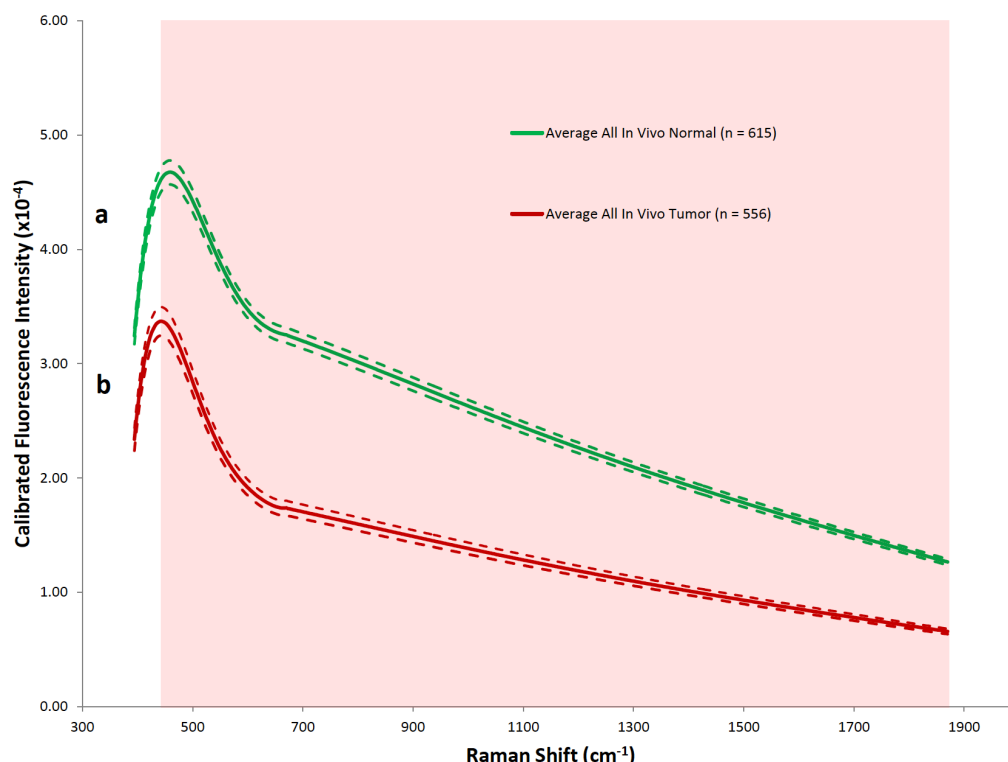


Figure 53. Average auto-fluorescence spectra from brain tissue derived from polynomial fits to the calibrated raw spectra.

From top to bottom the spectra are: a) average of all normal spectra ($n = 615$), and b) average of all tumour spectra ($n = 556$). Dashed lines are the $\pm$ uncertainties on the averages. Shaded wavebands are ranges over which a non-parametric statistic (Mann-Whitney U) indicates a highly significant ($p < 10^{-6}$) difference between the tumour and normal averages.

Although these average spectra were essentially featureless apart from a broad band from 394 – 670 cm^{-1} , there were statistically significant differences in intensities across virtually the whole measured range. Closer inspection of the average spectra also indicates the width of the broad band from 394 – 670 cm^{-1} is on average smaller for spectra in the all tumour pathology class. Normalizing each of the 1171 spectra by the area under the curve (prior to averaging by pathology) revealed emission ranges with different trends (Figure 54).

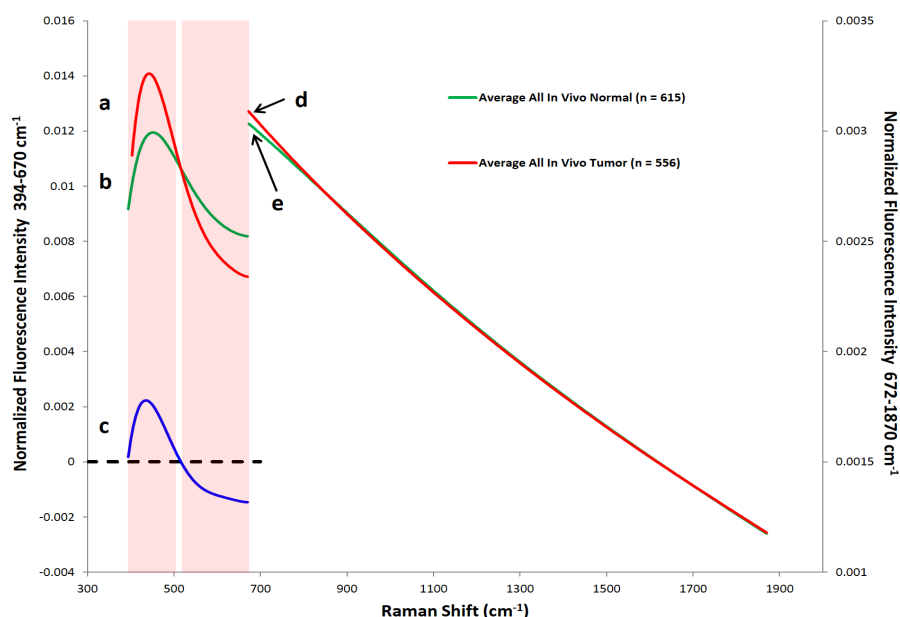


Figure 54. Average normalized auto-fluorescence spectra from brain tissue derived from polynomial fits to the calibrated raw spectra.

Spectra were normalized by the area under the curve in two ranges: 394 – 670 cm^{-1} and 672 – 1870 cm^{-1} . Spectra are plot on two different intensity scales for clarity. From top to bottom the spectra in the 394 – 670 cm^{-1} range are: a) average of all tumour spectra ($n = 556$), b) average of all normal spectra ($n = 615$), and c) difference between the tumour and normal averages. In the 672 – 1870 cm^{-1} range, spectra are identified by arrows: d) tumour and e) normal. Dashed lines are $\pm$ uncertainties on the averages. Shaded wavebands are ranges over which a non-parametric statistic (Mann-Whitney U) indicates a highly significant ($p < 10^{-6}$) difference between the tumour and normal averages.

From 394 – 670 cm^{-1} , the average normalized spectra show statistically significant differences between the tumour and normal averages. This result is consistent with having fluorescing molecules or processes at normal sites that are on average different to those at tumour sites. This would explain the change in the spectral band shape seen in Figure 53. In contrast for the range 672 – 1870 cm^{-1} the normalized spectra showed no statistically significant differences between the tumour and normal averages. This result is consistent with having the same dominant fluorescing molecule or process for all measurement sites.

A multivariate statistical analysis was performed of the 1171 auto-fluorescence spectra using PLS with waveband selection. The result showed some separation of the all tumour and all normal pathology classes (see Figures 55,57 and the inset of Figure 56), although the separation was not as great as that obtained with the Raman spectra. The area under the ROC curve (Figure 56) as a percentage of the total was 85% (95% CI of 83-87%), and a nominal sensitivity of 85% (95% CI of 82-88%) is obtained for a specificity of 69% (95% CI of 65-72%).

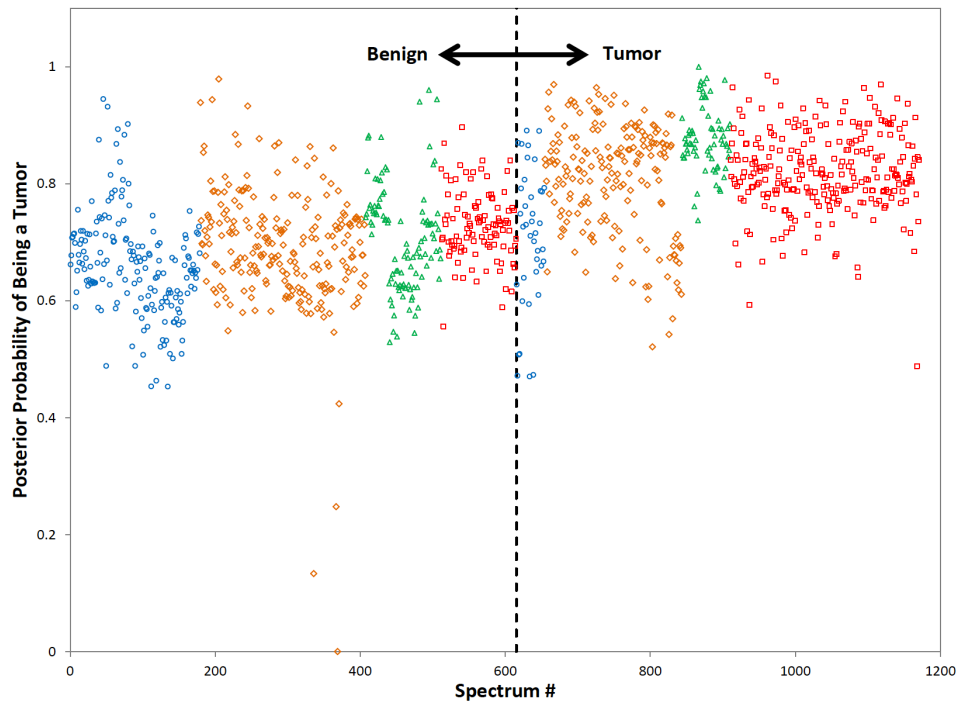


Figure 55. Posterior probability plots obtained from the analyses of 1171 auto-fluorescence spectra using PLS with waveband selection.

Each symbol represents the result of one spectrum and the different symbol shapes represent data obtained from patients with the following WHO grade tumours: $\circ$ = G2, $\diamond$ = G3 and $\triangle$ = G4 (GBM). $\square$ = spectra taken from patients with other tumour types.

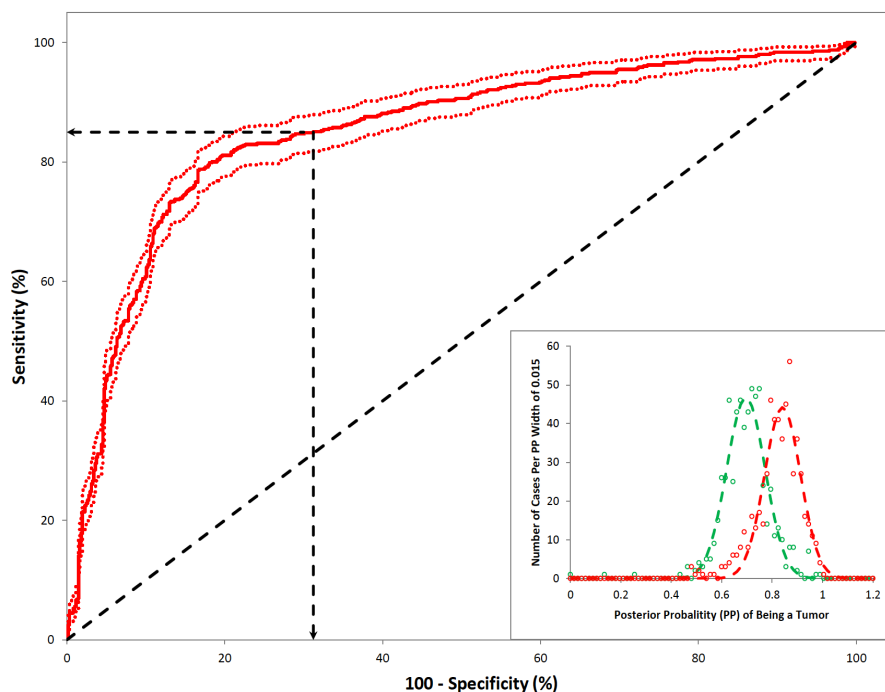


Figure 56. The ROC curve (continuous curve) and 95% CIs (dashed curves) derived from the posterior probabilities determined by PLS with waveband selection analyses of 1171 auto-fluorescence spectra.

Diagonal dashed line is the random guess, and the horizontal and vertical lines

intercept the ROC curve at a sensitivity of 85% (95% CI of 82-88%), a specificity of 69% (95% CI of 65-72%). The area under the curve as a percentage of the total is 85% (95% CI of 83-87%). Inset: frequency distribution in posterior probability values fitted with Gaussian models (Green circle= normal and red circle= tumour)

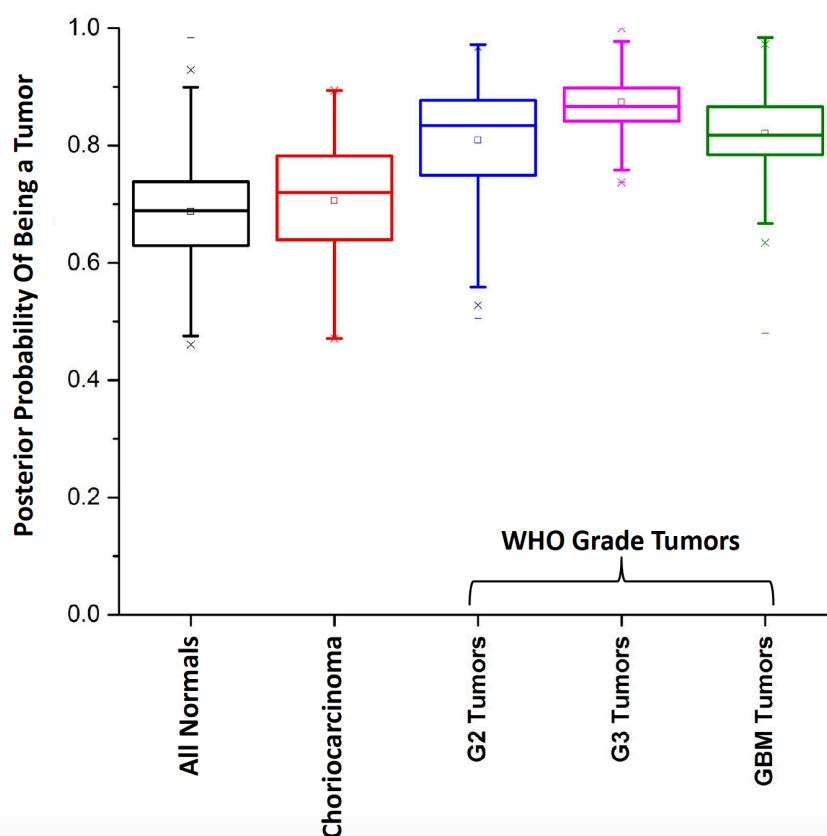


Figure 57. Boxplot representations of the posterior probability distributions as determined by PLS with waveband selection analyses of 1171 auto-fluorescence spectra.

The lower bound on each box shows the 25th percentile, where the upper bound shows the 75th percentile, meaning 25% and 75% of the data points are below these two bounds respectively. The line between these two bounds is the median. The whiskers on the plot show the 5th and 95th percentile, meaning 5% and 95% of the data are below these whiskers respectively. Data found outside the 5–95th percentile whiskers are outliers, as shown by separate data points.

A multivariate statistical analyses on 1171 combined Raman and auto-fluorescence spectra was then performed. Again, there was a clear separation of the all tumour and all normal pathology classes and a trend towards greater separation for later stage cancers (figures 58,59 and the inset of Figure 60). The area under the ROC curve, as a percentage of the total, increased to 93% (95% CI of 92-94%) compared to the analyses on the Raman or auto-fluorescence spectra alone (Figure 60). Correspondingly the prediction accuracy also increased where a sensitivity of 85% (95% CI of 82-88%), was now obtained for an increased specificity of 87% (95% CI of 84-90%). Again, out of the three analyses methods used, PLS with selected wavebands produced marginally better results than the PLS without waveband selection, or the boosted trees method (see Table 9). The prediction accuracy of the

953 spectra obtained from patients with WHO grade tumours also increased to a sensitivity of 86% with a specificity of 86% (Table 9). See Table 10 for a breakdown of results for different tumour subclasses.

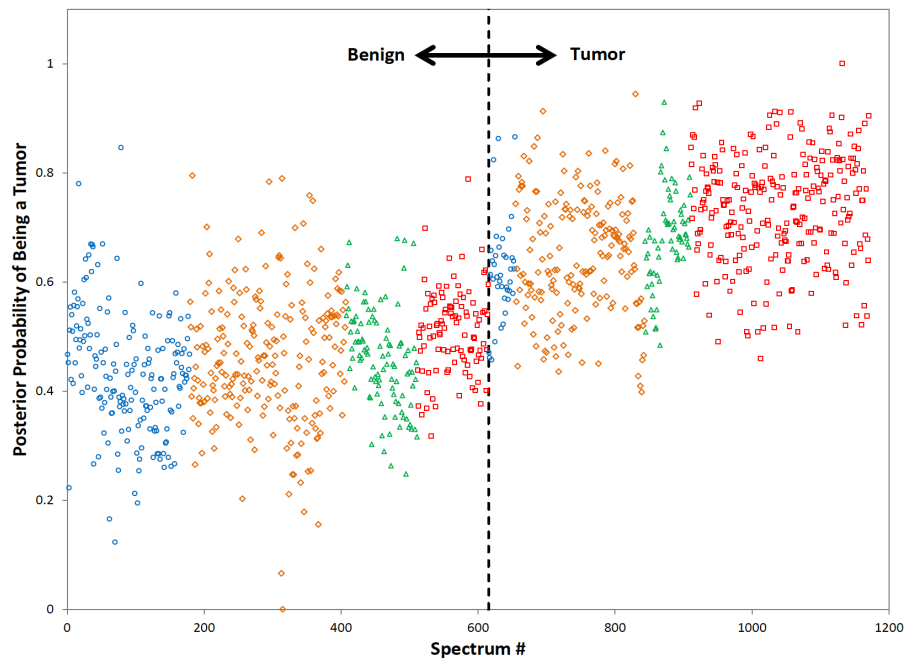


Figure 58. Posterior probability plots obtained from the analyses of 1171 combined Raman and auto-fluorescence spectra using PLS with waveband selection.

Each symbol represents the result of one spectrum and the different symbol shapes represent data obtained from patients with the following WHO grade tumours: $\circ$ = G2, $\diamond$ = G3, and $\triangle$ = G4 (GBM). $\square$ = spectra taken from patients with other tumour types.

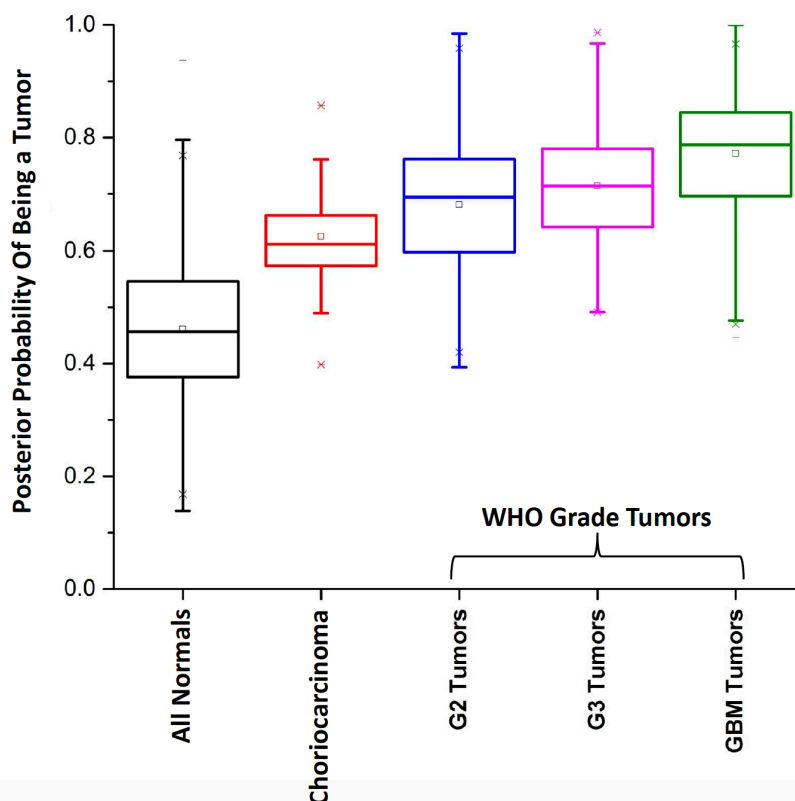


Figure 59. Boxplot representations of the posterior probability distributions as determined by PLS with waveband selection analyses of 1171 combined Raman and auto-fluorescence spectra.

The lower bound on each box shows the 25th percentile, where the upper bound shows the 75th percentile, meaning 25% and 75% of the data points are below these two bounds respectively. The line between these two bounds is the median. The whiskers on the plot show the 5th and 95th percentile, meaning 5% and 95% of the data are below these whiskers respectively. Data found outside the 5–95th percentile whiskers are outliers, as shown by separate data points.

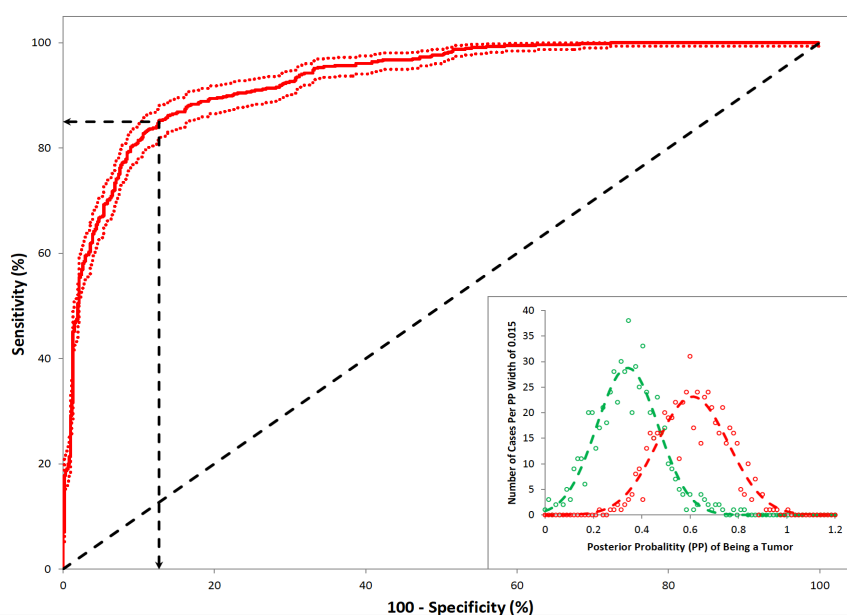


Figure 60. The ROC curve (continuous curve) and 95% CIs (dashed curves) derived from the posterior probabilities determined by PLS with waveband

selection analyses of 1171 combined Raman and auto-fluorescence spectra.

Diagonal dashed line is the random guess, and the horizontal and vertical lines intercept the ROC curve at a sensitivity of 85% (95% CI of 82-88%), a specificity of 87% (95% CI of 84-90%). The area under the curve as a percentage of the total is 93% (95% CI of 92-94%). Inset: frequency distribution in posterior probability values fitted with Gaussian models (Green circle= normal and red circle= tumour)

Results of the Statistical Analyses on 1171 Combined Raman and Fluorescence Spectra					
WHO Grade	Tumor Type	# Of Patients	Specificity (%)	Sensitivity (%)	Accuracy (%)
G2	G2 Astro	5	86	73	80
	G2 Oligo	2	79	89	84
	G2 Oligo/GBM	1	86	0	64
	G2 Type Totals	8	84	75	80
G3	G3	1	84	65	76
	G3 Oligo	1	NA	100	100
	G3 Oligo (GBM)	1	100	100	100
	G3 Oligoastro	1	89	NA	89
	G3 Type Totals	4	91	88	90
G4	GBM	8	84	93	90
G2,G3 and G4 Type Totals		20	86	86	86
Other	Hemangioma 4C	1	96	NA	96
	Lung Met	1	100	NA	100
	Meningioma	1	97	NA	97
	Radionecrosis	1	92	NA	92
	Choriocarcinoma	1	74	77	75
	Grade Totals	5	91	77	89
All Grades	All Types	25	87	85	86

Table 10. Classification of different tumour subclasses based on PLS analyses with waveband section.

Classification was based on training and test spectra drawn from the full data set of 1171 normalized and combined Raman and Fluorescence spectra acquired from patients with all tumour classes.

The multivariate statistical analyses were repeated on the Raman, auto-fluorescence and combined Raman and auto-fluorescence spectra, but only spectra from patients with confirmed glioma (G2, G3 and G4 GBM) were used for training and testing. The results again showed a clear separation of the two pathology classes (all normal and all tumour) and a trend towards greater separation for later stage cancers for the three spectral subsets analysed. PLS with selected wavebands produced marginally better results than the PLS without waveband selection, or the boosted trees method (see table 11). Figures 61, 62 and the inset of Figure 63 show the distribution in posterior probability values obtained with the PLS with selected wavebands method. The area under the ROC curve (Figure 63) as a percentage of the total area was 95% (95% CI of 93-96%); this was larger than that obtained with the same analyses method on all 1171 combined Raman or auto-fluorescence spectra. Correspondingly the prediction accuracy also increased where a nominal sensitivity of 85% (95% CI of 82-88%) was now obtained for an increased specificity of 92% (95% CI of 88-94%). See Table 12 for a breakdown of results for different WHO grades.

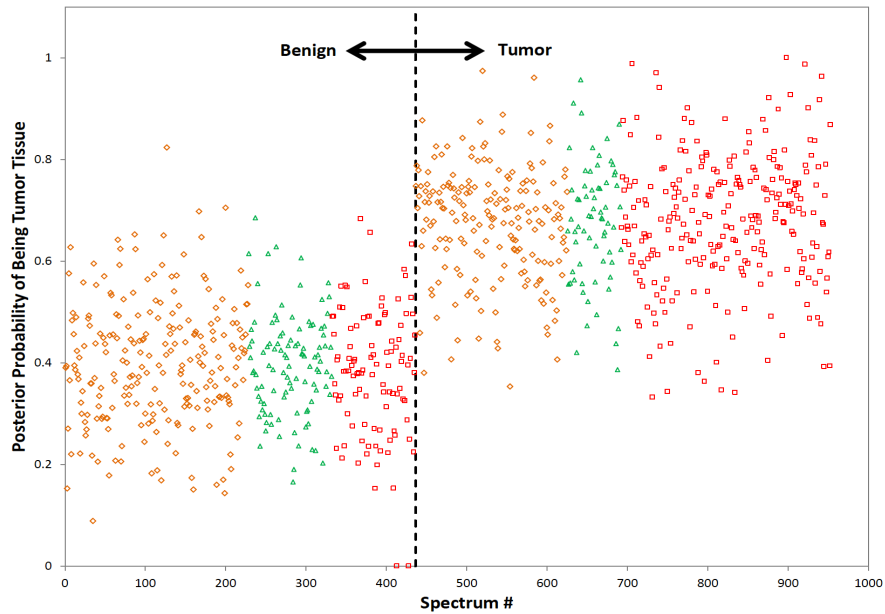


Figure 61. Posterior probability plots from 953 combined Raman and auto-fluorescence spectra taken from gliomas.

The PLS with waveband selection method was used. Each symbol represents the result of one spectrum and the different symbol shapes represent data obtained from patients with the following WHO grade tumours: $\diamond$ = G2, $\triangle$ = G3, and $\square$ = G4 (GBM).

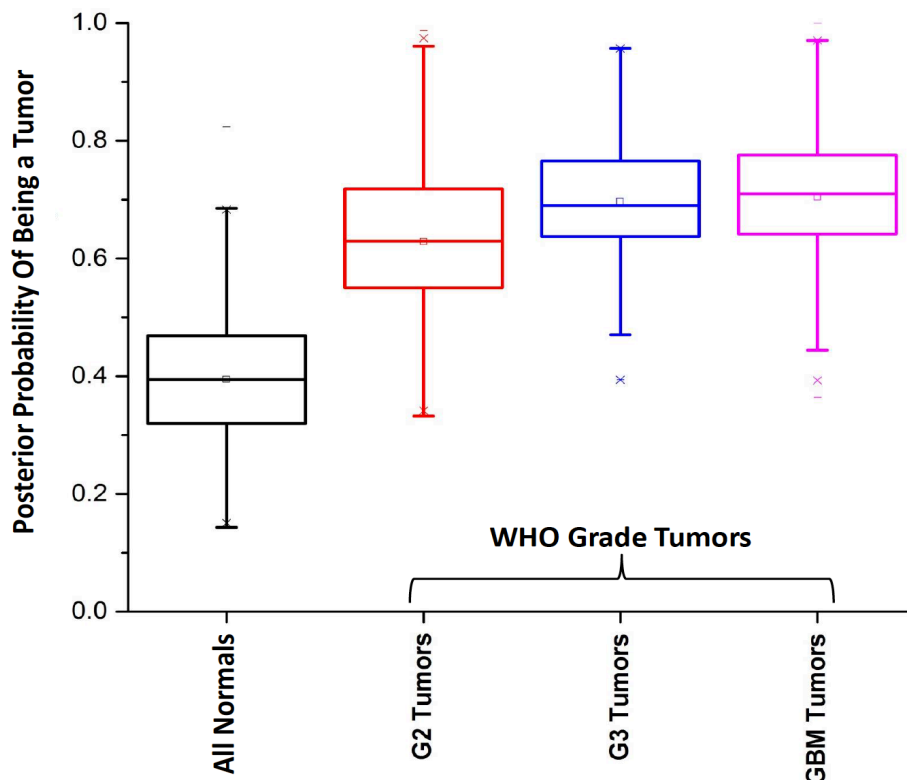


Figure 62. Box plot representations of the posterior probability distributions as determined by PLS with waveband selection from 953 combined Raman and auto-fluorescence spectra from gliomas.

The lower bound on each box shows the 25th percentile, where the upper bound

shows the 75th percentile, meaning 25% and 75% of the data points are below these two bounds respectively. The line between these two bounds is the median. The whiskers on the plot show the 5th and 95th percentile, meaning 5% and 95% of the data are below these whiskers respectively. Data found outside the 5–95th percentile whiskers are outliers, as shown by separate data points. Gliomas exhibit a clear difference from normal tissue.

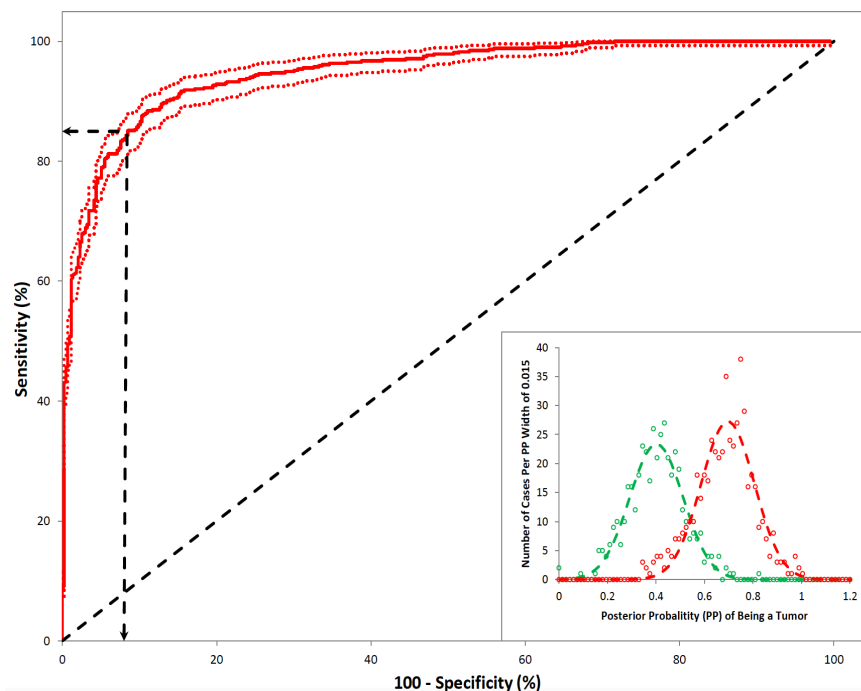


Figure 63. The ROC curve (continuous curve) and 95% CIs (dashed curves) derived from the posterior probabilities determined by PLS with waveband selection analyses of 953 combined Raman and auto-fluorescence spectra taken from patients with gliomas.

Diagonal dashed line indicates a classifier which is as good as a random guess and the horizontal and vertical lines intercept the ROC curve at a sensitivity of 85% (95% CI of 82-88%), a specificity of 92% (95% CI of 88-94%). The area under the curve as a percentage of the total is 95% (95% CI of 93-96%). Inset: frequency distribution in posterior probability values fitted with Gaussian models (Green circle= normal and red circle= tumour)

Results of the Statistical Analyses on 953 Combined Raman and Fluorescence Spectra					
WHO Grade	Tumor Type	# Of Patients	Specificity (%)	Sensitivity (%)	Accuracy (%)
G2	G2 Astro	5	92	71	83
	G2 Oligo	2	83	85	84
	G2 Oligo/GBM	1	97	33	81
	G2 Type Totals	8	90	74	83
G3	G3	1	94	74	85
	G3 Oligo	1	NA	100	100
	G3 Oligo (GBM)	1	100	100	100
	G3 Oligoastro	1	89	NA	89
	G3 Type Totals	4	94	91	93
G4	GBM	8	91	92	92
G2,G3 and G4 Type Totals		20	92	85	88

Table 12. Classification of different tumour subclasses based on PLS analyses with waveband section.

Classification was based on training and test spectra drawn from a data subset of 953 normalized and combined Raman and Fluorescence spectra acquired from patients with WHO graded tumours (G2, G3 and G4).

Analyses Results		Test Data	G2, G3 and G4 Cases Only	
Training Data	Method	AUC (%)	Spec. (%)	Sens. (%)
Raman From G2, G3 or G4 Patients Only	Boosted Trees	90 (88 to 92)	69 (64 to 73)	85 (82 to 88)
	Partial Least Squares Full	89 (87 to 90)	78 (73 to 81)	85 (82 to 88)
	Partial Least Squares Select	92 (90 to 93)	84 (80 to 87)	85 (82 to 88)
Fluorescence From G2, G3 or G4 Patients Only	Boosted Trees	89 (87 to 91)	61 (56 to 65)	85 (82 to 88)
	Partial Least Squares Full	86 (84 to 88)	72 (67 to 76)	85 (82 to 88)
	Partial Least Squares Select	86 (84 to 88)	72 (67 to 76)	85 (82 to 88)
Raman and Fluorescence From G2, G3 or G4 Patients Only	Boosted Trees	94 (92 to 95)	87 (83 to 90)	85 (82 to 88)
	Partial Least Squares Full	94 (92 to 95)	89 (85 to 92)	85 (82 to 88)
	Partial Least Squares Select	95 (93 to 96)	92 (88 to 94)	85 (82 to 88)

Table 13. Classification of spectra acquired from patients with different tumour subclasses based on different multivariate methods.

Percentages in parentheses are the 95% confidence intervals. Classification was based on training and test spectra drawn from a data subset of 953 normalized and combined Raman and Fluorescence spectra acquired from patients with WHO graded tumours (G2, G3 and G4). Two classifications were performed, a) using Raman spectra only and b) using combined Raman and Fluorescence spectra.

I used a range of other machine learning methods on the *in vivo* data to test for robustness of classifiers. The data was run through 2 data analysis packages: Rapidminer (Rapidminder Ltd, London) and Rattle (Graham Williams, Togaware). Both packages were able to apply the whole range of modelling techniques including SVM, RF, BT and Neural Nets. All classifiers performed well, with the best classifiers having accuracy in the high 80% to 90% range. The confusion matrices, ROC AUC's and accuracies are shown below for the best performing models.

Rapidminer model

Input: Rattle TvsN.xls (all *in vivo* data)

Data split: 80% training, 20% testing

Gradient boosted trees, SVM, Neural Net, Random Forest used. Best results obtained with boosted trees. Seems to over classify normal into tumour.

Gradient Boosted Trees

PerformanceVector:

accuracy: 80.93%

ConfusionMatrix:

True:	Normal	Tumour
-------	--------	--------

Normal:	92	21
---------	----	----

Tumour:	20	82
---------	----	----

precision: 80.39% (positive class: Tumour)

ConfusionMatrix:

True:	Normal	Tumour
-------	--------	--------

Normal:	92	21
---------	----	----

Tumour:	20	82
---------	----	----

recall: 79.61% (positive class: Tumour)

ConfusionMatrix:

True:	Normal	Tumour
-------	--------	--------

Normal:	92	21
---------	----	----

Tumour:	20	82
---------	----	----

AUC (optimistic): 0.898 (positive class: Tumour)

AUC: 0.898 (positive class: Tumour)

AUC (pessimistic): 0.898 (positive class: Tumour)

Rattle model

Input: Rattle TvsN.xls (all *in vivo* data)

Data split: 80% training, 20% testing

Gradient boosted trees, SVM, Neural Net, Random Forest used. Best results:

89% accuracy with Random Forrest

Error matrix for the Random Forest model on Rattle TvsN.csv [test] (counts):

Predicted		
Actual	Normal	Tumour
Normal	96	11
Tumour	13	96

Overall error: 11%, Averaged class error: 11%

The associated ROC curve shows a good 0.95 AUC as shown in Figure 64.

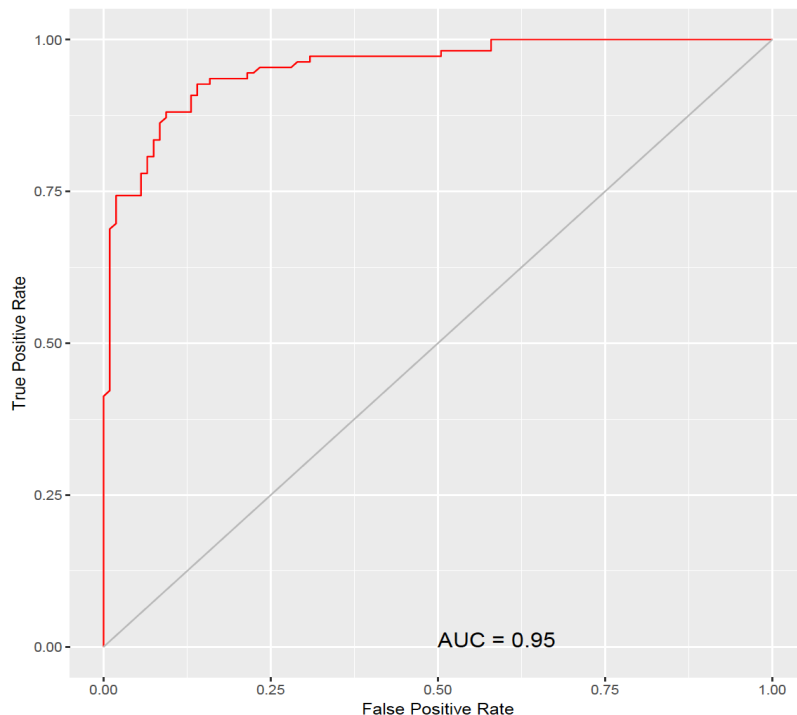


Figure 64. ROC curve of the Random Forrest model classifying tumour vs normal spectra from all 1171 spectra.

80% training and 20% testing data split used. The AUC is 0.95 indicating good classifier performance in the testing data set.

The data analysis was repeated for biopsy tissue only. This revealed much weaker models due to the data being unbalanced (it is obviously unethical to biopsy tissue during an operation for a brain tumour) and much fewer spectra to train models on. The key classification errors involved overclassifying normal spectra as tumour spectra as the training dataset was biased with much higher numbers of tumour spectra than normal. The best performing classifier was the Ada Boosted Trees, with other classifiers being similar or inferior as follows:

Rattle: Biopsy only ALL.csv, 70/30 partition.

Error matrix for the Ada Boost model on Biopsy only ALL.csv [test] (counts):

Actual	Predicted	
	Normal	Tumour
Normal	10	8
Tumour	0	52

Error matrix for the Ada Boost model on Biopsy only ALL.csv [test] (proportions):

Actual	Predicted	
	Normal	Tumour Error

Normal	0.14	0.11	0.44
Tumour	0.00	0.74	0.00

Overall error: 11%, Averaged class error: 22%

Rattle timestamp: 2017-06-18 00:37:51 Babar Vaqas

=====

=====

Error matrix for the Random Forest model on Biopsy only ALL.csv [test] (counts):

		Predicted	
Actual	Normal	Tumour	
Normal	9	9	
Tumour	0	52	

Error matrix for the Random Forest model on Biopsy only ALL.csv [test]
(proportions):

		Predicted		
Actual	Normal	Tumour	Error	
Normal	0.13	0.13	0.5	
Tumour	0.00	0.74	0.0	

Overall error: 13%, Averaged class error: 25%

Rattle timestamp: 2017-06-18 00:37:52 Babar Vaqas

=====

=====

Error matrix for the SVM model on Biopsy only ALL.csv [test] (counts):

		Predicted	
Actual	Normal	Tumour	
Normal	10	8	
Tumour	0	52	

Error matrix for the SVM model on Biopsy only ALL.csv [test] (proportions):

		Predicted		
Actual	Normal	Tumour	Error	
Normal	0.14	0.11	0.44	
Tumour	0.00	0.74	0.00	

Overall error: 11%, Averaged class error: 22%

Rattle timestamp: 2017-06-18 00:37:53 Babar Vaqas

=====

=====

To address the concerns about bias in the training data set, a matched training set was created with equal numbers of normal and tumour spectra from biopsy only

data. Tumour spectra were chosen at random, thereby eliminating the imbalance between normal and tumour spectra. Another approach was to define tumour samples as those not with a report of infiltration: this reduced the number of tumour samples in the training dataset reducing the imbalance between normal and tumour samples.

Reducing the imbalance between normal and tumour samples resulted in only 1 well performing model: Random Forest as shown below:

Rattle: reduced imbalance dataset
 68 Normal 163 tumour spectra
 60/40 data split

Error matrix for the Random Forest model on Biopsy only deleted tumours.csv [test] (counts):

	Predicted	
Actual	Normal	Tumour
Normal	10	2
Tumour	0	19

Error matrix for the Random Forest model on Biopsy only deleted tumours.csv [test] (proportions):

	Predicted		
Actual	Normal	Tumour	Error
Normal	0.32	0.06	0.17
Tumour	0.00	0.61	0.00

Overall error: 6%, Averaged class error: 8%

Rattle timestamp: 2017-06-18 11:28:51 Babar Vaqas

=====

Error matrix for the SVM model on Biopsy only deleted tumours.csv [test] (counts):

	Predicted	
Actual	Normal	Tumour
Normal	10	2
Tumour	0	19

Error matrix for the SVM model on Biopsy only deleted tumours.csv [test] (proportions):

	Predicted		
Actual	Normal	Tumour	Error
Normal	0.32	0.06	0.17
Tumour	0.00	0.61	0.00

Overall error: 6%, Averaged class error: 8%

Rattle timestamp: 2017-06-18 11:28:52 Babar Vaqas

=====

Rapidminer

80/20 split, Boosted Trees

Rapid miner gives good results with high sensitivity for Tumour at 93% sensitivity and 88% specificity.

To eliminate the imbalance between tumour and normal spectra 94 normal and 94 randomly chosen tumour spectra were matched. I obtained a higher classification accuracy using RF and SVM as shown below:

Rattle

94 Normal, 94 Tumour

80/20 training/testing data partition

Error matrix for the Random Forest model on Biopsy only tumour converted to normal and 43 Tumour 2 deleted.csv [test] (counts):

	Predicted	
Actual	Normal	Tumour
Normal	20	3
Tumour	0	15

Error matrix for the Random Forest model on Biopsy only tumour converted to normal and 43 Tumour 2 deleted.csv [test] (proportions):

	Predicted		
Actual	Normal	Tumour	Error
Normal	0.53	0.08	0.13
Tumour	0.00	0.39	0.00

Overall error: 8%, Averaged class error: 6%

Rattle timestamp: 2017-06-18 13:01:21 Babar Vaqas

=====

Error matrix for the SVM model on Biopsy only tumour converted to normal and 43 Tumour 2 deleted.csv [test] (counts):

	Predicted	
Actual	Normal	Tumour
Normal	20	3
Tumour	0	15

Error matrix for the SVM model on Biopsy only tumour converted to normal and 43 Tumour 2 deleted.csv [test] (proportions):

	Predicted		
Actual	Normal	Tumour	Error
Normal	0.53	0.08	0.13
Tumour	0.00	0.39	0.00

Overall error: 8%, Averaged class error: 6%

Rattle timestamp: 2017-06-18 13:01:22 Babar Vaqas

=====

The RF feature importance was as follows:

	Normal	Tumour	MeanAccuracy	MeanDecreaseGini
X1592.516799	3.96	3.24	4.55	0.62
X877.6081193	3.23	2.77	4.39	0.46
X1628.423302	3.16	3.03	4.05	0.52
X875.0007308	3.57	3.34	4.03	0.62
X1626.187168	3.58	3.33	3.98	0.51
X1606.013895	2.32	3.62	3.85	0.39
X911.3830789	3.72	2.57	3.83	0.42
X880.2141731	3.29	3.19	3.82	0.42
X1632.892374	1.79	4.12	3.80	0.48
X748.2565666	3.72	2.40	3.77	0.24
X1422.811897	3.58	3.14	3.76	0.41
X1599.270186	2.55	3.39	3.66	0.36
X1594.769004	2.97	2.86	3.65	0.40
X1612.747949	2.86	3.12	3.60	0.37
X1623.949968	1.96	3.27	3.58	0.34
X1425.145918	2.73	2.73	3.52	0.45
X1603.767066	1.23	3.91	3.52	0.43

To investigate the generalisability of the biopsy model created, I used the whole biopsy data as training set to create a model for testing on the non-biopsy data. This resulted in a RF classifier with an 83% accuracy for identifying normal tissue but very poor accuracy for tumour tissue. The SVM model performed poorly for normal tissue identification (92% error rate) but classified tumour tissue with 92% accuracy. The model error matrices are below:

Rattle

Training set: biopsy spectra

Test set: all non-biopsy spectra

Error matrix for the Random Forest model on Rattle TvsN no overlap biopsy only deleted or converted ExP30.csv (counts):

	Predicted	
Actual	Normal	Tumour
Normal	370	74

Tumour 285 145

Error matrix for the Random Forest model on Rattle TvsN no overlap biopsy only deleted or converted ExP30.csv (proportions):

	Predicted		
Actual	Normal	Tumour	Error
Normal	0.42	0.08	0.17
Tumour	0.33	0.17	0.66

Overall error: 41%, Averaged class error: 42%

83% accurate for Normal tissue

Error matrix for the SVM model on Rattle TvsN no overlap biopsy only deleted or converted ExP30.csv (counts):

	Predicted	
Actual	Normal	Tumour
Normal	35	409
Tumour	34	396

Error matrix for the SVM model on Rattle TvsN no overlap biopsy only deleted or converted ExP30.csv (proportions):

	Predicted		
Actual	Normal	Tumour	Error
Normal	0.04	0.47	0.92
Tumour	0.04	0.45	0.08

Overall error: 51%, Averaged class error: 50%

92% accurate for tumour tissue

Rattle timestamp: 2017-06-18 13:16:50 Babar Vaqas

=====

The next and final step in the 'tumour vs normal' analysis was to combine *ex vivo* and *in vivo* data to maximise the dataset size and then to see if accurate classification models can be produced from this data. As shown in the previous section, both of these data types cannot be distinguished easily from each other using machine learning, suggesting that both can be combined. This creates a robust training and testing dataset consisting of 2226 spectra from 29 patients.

The Cubic SVM model I created based on this data reaches an impressive 88.1% accuracy on 8 fold cross validation (278 spectra in the test set) as shown in Figure 65.

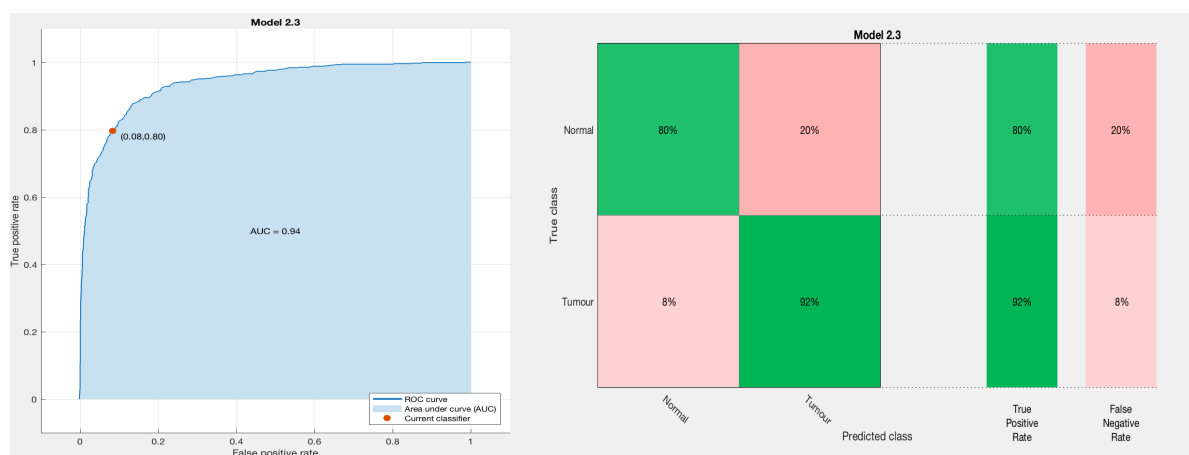


Figure 65. Confusion matrix and ROC curve of the cubic SVM classifier for 2226 *in vivo* and *ex vivo* spectra.

The overall accuracy of this machine learning method using the combined dataset is 88.1%. 80% of normal spectra were identified and 92% of tumour spectra were identified successfully. The AUC is an impressive 0.94 suggesting good performance of this classifier.

80% of normal spectra were successfully identified, with 92% of tumour spectra successfully identified. The AUC for the ROC curve was 0.94 as illustrated in Figure 65 suggesting excellent performance of the model.

4.7.5.4. Summary

The results I obtained in this section show that tumour and normal tissue can successfully be identified using machine learning on the basis of Raman spectroscopy. The accuracy I achieved is increased when adding in information from autofluorescence. The best results were achieved using the PLS-DA method on Raman and autofluorescence data resulting in 85% sensitivity and 87% specificity, with an AUC of the ROC curve of 0.92. Interestingly, classifying on autofluorescence alone is possible and this yields inferior classification accuracy with a lower specificity of 61-72%, suggesting that some diagnostic information is contained in fluorescence data.

When dealing with the smaller biopsy only dataset, I found that models created with RF were the best for dealing with imbalanced data as good classification ability was shown on cross validation with 20% of data which was held out. However, when used to classify non-biopsy spectra, the biopsy only model has limited generalisability with RF being better for detecting normal spectra with 83% sensitivity and SVM being better for detecting tumour spectra with 92% sensitivity. Using a balanced training dataset with equal numbers of tumour and normal spectra I produced improved models with accuracy reaching 87% for normal spectra and 100% for tumour spectra. This was however based on cross validation with a small test set.

Using the complete dataset (*in vivo* and *ex vivo* data combined) resulted in a cubic

SVM model with which I achieved 88.1% accuracy. This is significant as the dataset consisted of 2226 spectra and the testing dataset was large at 278 spectra, indicating a robust algorithm for detecting tumour tissue.

4.7.6. Tumour Density

4.7.6.1. Introduction

As shown in the previous section tumour tissue detection is possible with high accuracy using PLS-DA and RF. The next question is whether there is a semi-quantitative or quantitative relationship between tumour density and the associated Raman spectrum. Tumour density is commented on clinically using a semi-quantitative scale, whereas counting tumour cells is not currently performed as part of histopathological reporting.

This chapter looks at the relationship between the Raman spectrum and pathologist reporting of tumour density. The aim is to train models able to predict the tumour density of a tissue sample from its Raman spectrum in order to aid the surgeon when determining the resection margin in diffuse disease such as glioblastoma.

For classification purposes, tumour density was classified as Normal tissue (zero density), infiltrated brain tissue (intermediate density) and tumour tissue (high density, solid tumour). This correlated well with the histopathological reports which allowed these states to be easily distinguished as they were commented on directly.

4.7.6.2. Multivariate analysis

I performed unsupervised analysis using PCA with and without fluorescence data to look for underlying structure in the data. Figure 66 shows poor clustering of data based on Raman signal alone with significant overlap between the 3 tumour density groups.

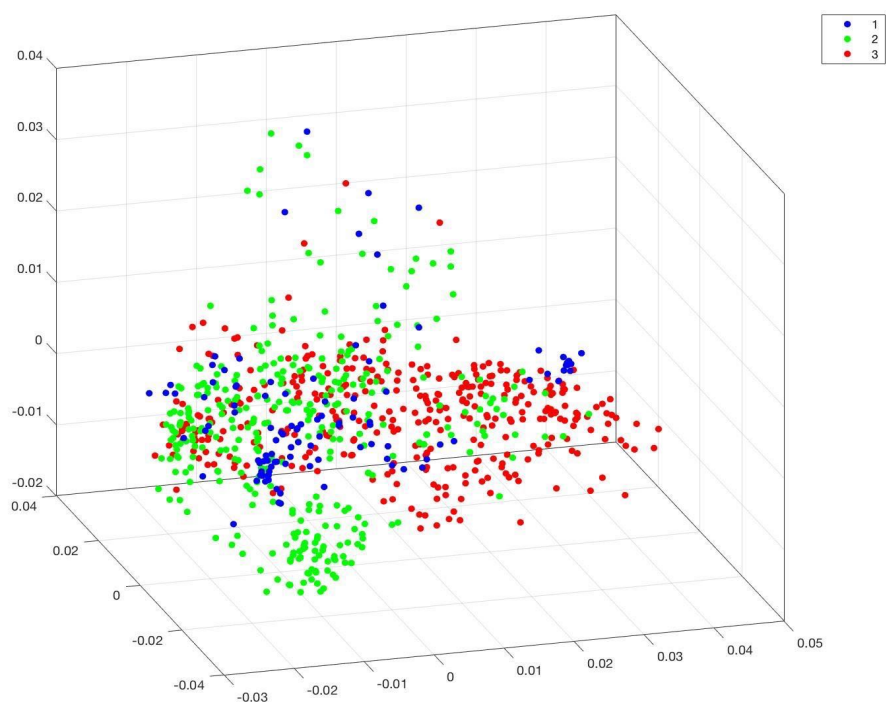


Figure 66. PCA plot of Raman spectra from biopsied tissue labelled for tissue density.

Top 3 principle components are shown. 1=normal, 2=infiltration and 3=tumour. There is poor clustering between classes with significant overlap.

Given the improved multivariate analyses in the tumour vs normal comparison in chapter 3.7.5 the data was processed using fluorescence signal as well as Raman signal. This improved the class separation on PCA as shown in Figure 67.

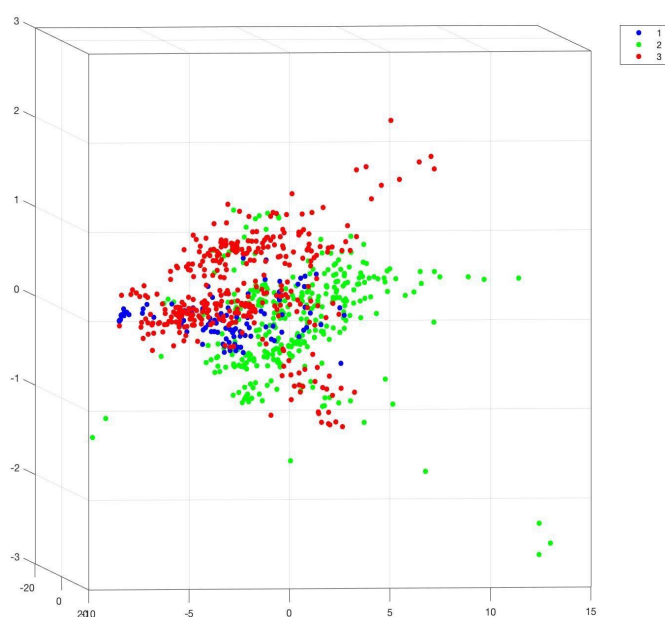


Figure 67. PCA plot of Raman spectra with fluorescence from biopsied tissue labelled for tissue density.

Top 3 principle components are shown. 1=normal, 2=infiltration and 3=tumour. There is improved clustering between classes with reduced overlap compared to Figure 66.

The improved clustering of data with fluorescence suggests supervised analysis should be successful in separating classes. LDA was performed on the Raman data with and without fluorescence as shown in Figure 68

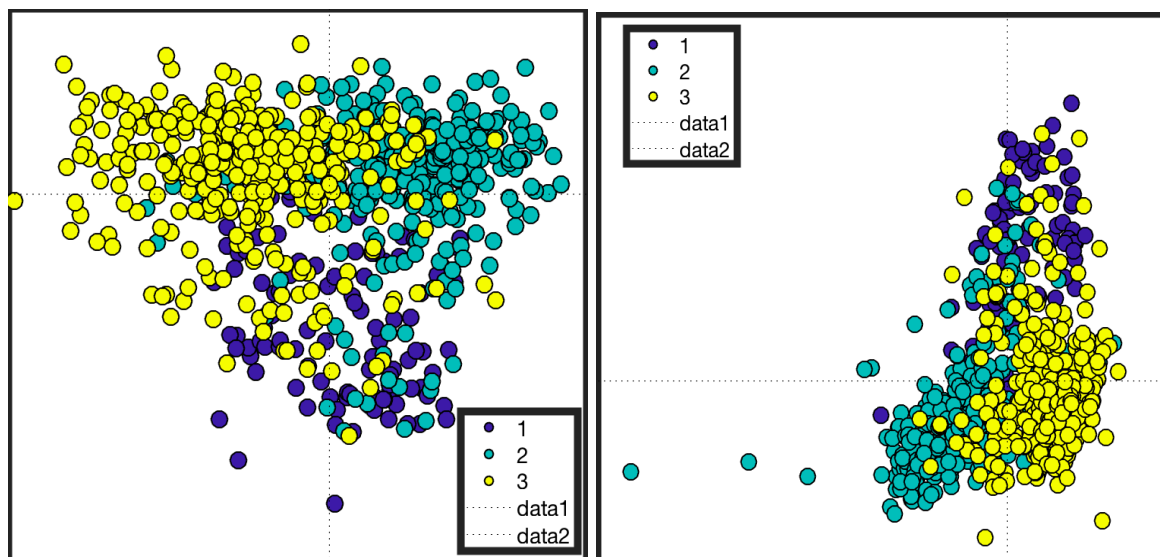


Figure 68. LDA plots of Raman spectra (left) and Raman and fluorescence spectra (right) for tumour density.

1=normal, 2=infiltration and 3=tumour. Tighter clustering of spectra but greater overlap is seen when fluorescence data is used together with Raman. Both LDA plots show good separation of classes.

I then constructed models using the 119 biopsy spectra which had tumour density information. An 80/20 training/test data split was used for validation. Analysis was performed on data with and without fluorescence signal. Without fluorescence signal the best performing models was SVM with 83% sensitivity and 82% specificity as below:

Error matrix for the SVM model on Rattle Tumour Density.csv [test] (counts):

	Predicted		
Actual	Infiltrated	Normal	Tumour
Infiltrated	21	2	4
Normal	10	91	10
Tumour	6	9	86

Error matrix for the SVM model on Rattle Tumour Density.csv [test] (proportions):

	Predicted			
Actual	Infiltrated	Normal	Tumour	Error

Infiltrated	0.09	0.01	0.02	0.22
Normal	0.04	0.38	0.04	0.18
Tumour	0.03	0.04	0.36	0.15

Overall error: 17%, Averaged class error: 18%

Rattle timestamp: 2018-04-03 15:05:55 Babar Vaqas

=====

Adding fluorescence to the dataset improved accuracy to 85% with RF as shown below:

Error matrix for the Random Forest model on Rattle Tumour Density Ram+Fluo.csv [test] (counts):

Actual	Predicted		
	Infiltrated	Normal	Tumour
Infiltrated	21	4	2
Normal	6	96	9
Tumour	3	12	86

Error matrix for the Random Forest model on Rattle Tumour Density Ram+Fluo.csv [test] (proportions):

Actual	Predicted			
	Infiltrated	Normal	Tumour	Error
Infiltrated	0.09	0.02	0.01	0.22
Normal	0.03	0.40	0.04	0.14
Tumour	0.01	0.05	0.36	0.15

Overall error: 15%, Averaged class error: 17%

Rattle timestamp: 2018-04-03 18:42:08 Babar Vaqas

=====

4.7.6.3. Summary

Supervised and unsupervised analyses showed the tendency of tumour density Raman data to cluster together. The predictive models I constructed performed well with the best accuracy obtained when using fluorescence data with Raman data achieving a sensitivity of 85% for determining tumour density in terms of normal (0 density), tumour (high density) or infiltrated tissue (intermediate density). Even without fluorescence models using Raman alone achieved 83% accuracy.

This suggests Raman spectroscopy can be used to determine tumour density in a semi-quantitative fashion. The next question which is connected to that of detecting

tumour density is whether Raman can be used to determine tumour margins during surgery.

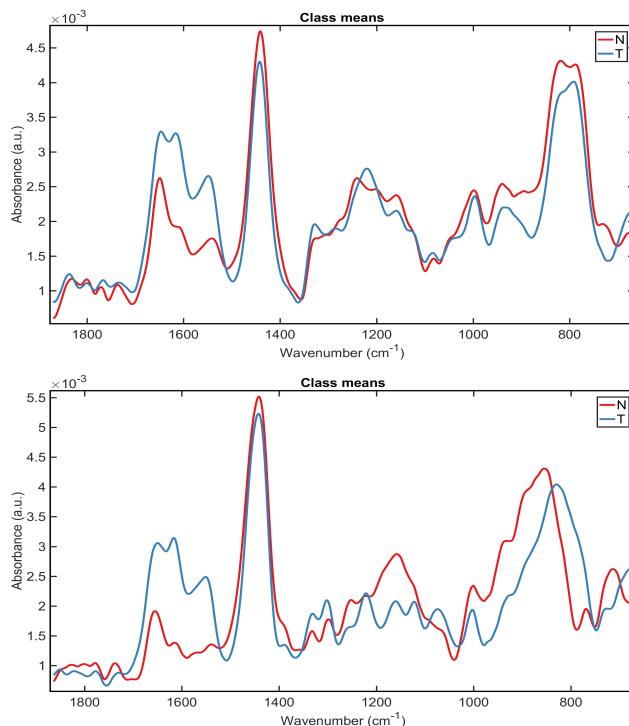
4.7.7. Tumour Grade

4.7.7.1. Introduction

Tumour grade is a clinic-pathological concept which is in part to do with biological aggression of the tumour, response to clinical management and the prognosis for the patient. Although an integration of multiple factors it is still in part influenced by the molecular biology of the tumour itself, so the hypothesis is that any molecular information gained from Raman spectroscopy may help classify tumour tissue into one of the 4 grades used to help determine treatment. This would of course be clinically significant in speeding up treatment times, producing streamlined patient pathways and selecting appropriate treatment as early as possible as discussed in section 2.4.

4.7.7.2. Average spectra

Grade 2, Grade 3 and WHO grade 4 tumours have uniquely different average spectra as shown in Figure 69 below.



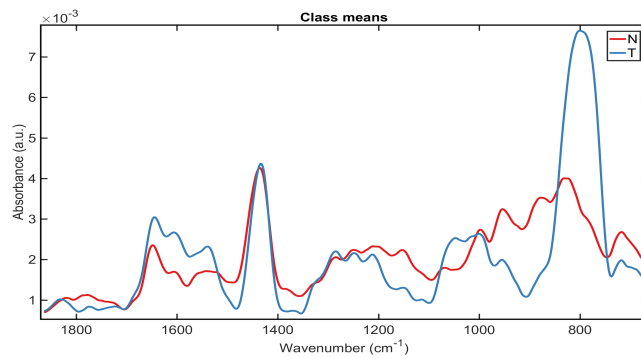


Figure 69. Average Raman spectra of the different tissue types included in the Raman study.

The tissue types have been grouped into grade 4 GBM (top), grade 2 astrocytoma (2nd from top), grade 2 oligodendroglioma (2nd from bottom). This illustrates the variation between tissue types and grades which may form the basis for classification models used to differentiate tissue types from one another.

In the case of grade 2 and grade 3 tumours, the histological diagnosis does not directly correspond to the tumour grade. Interestingly, where in the WHO classification it can be quite difficult to distinguish between grade 2 and 3 (to the extent that molecular markers have only recently been developed and introduced to help with this) the Raman spectral averages are very different as shown in Figure 70.

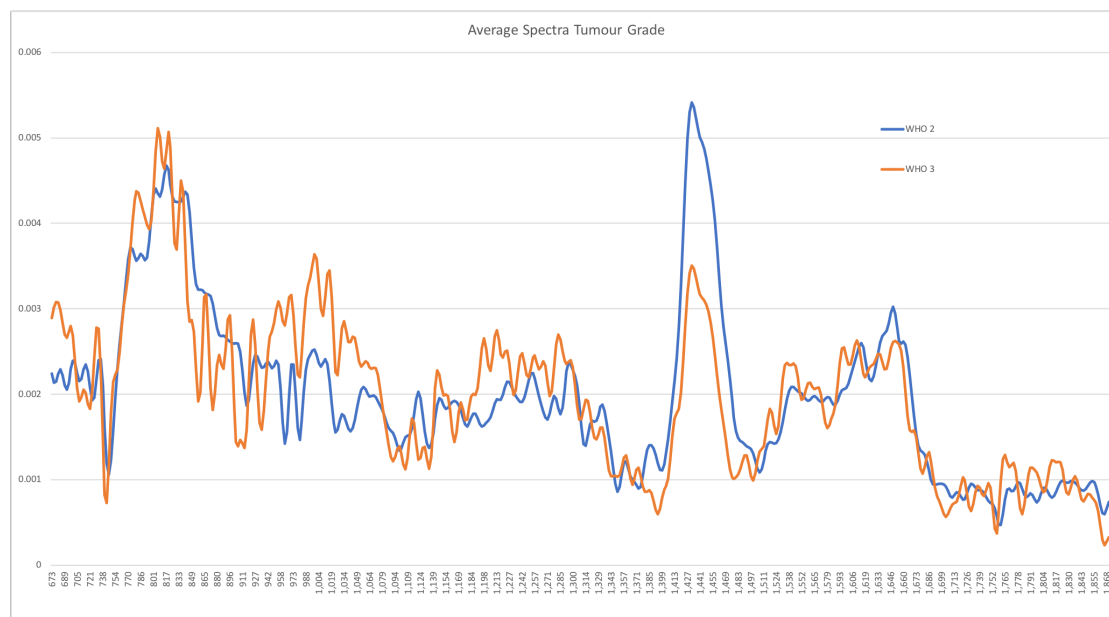


Figure 70. Average Raman spectra of WHO grade 2 and grade 3 tumours compared.

The most significant areas of difference are 1400-1490 and 940-1080 regions.

4.7.7.3. Multivariate analysis

I performed PCA analysis to see if the data segregated on unsupervised analysis alone. When performed on biopsy proven histology clustering was observed in the

168 spectra analysed as shown in Figure 71.

PCA WHO Grades: Biopsy

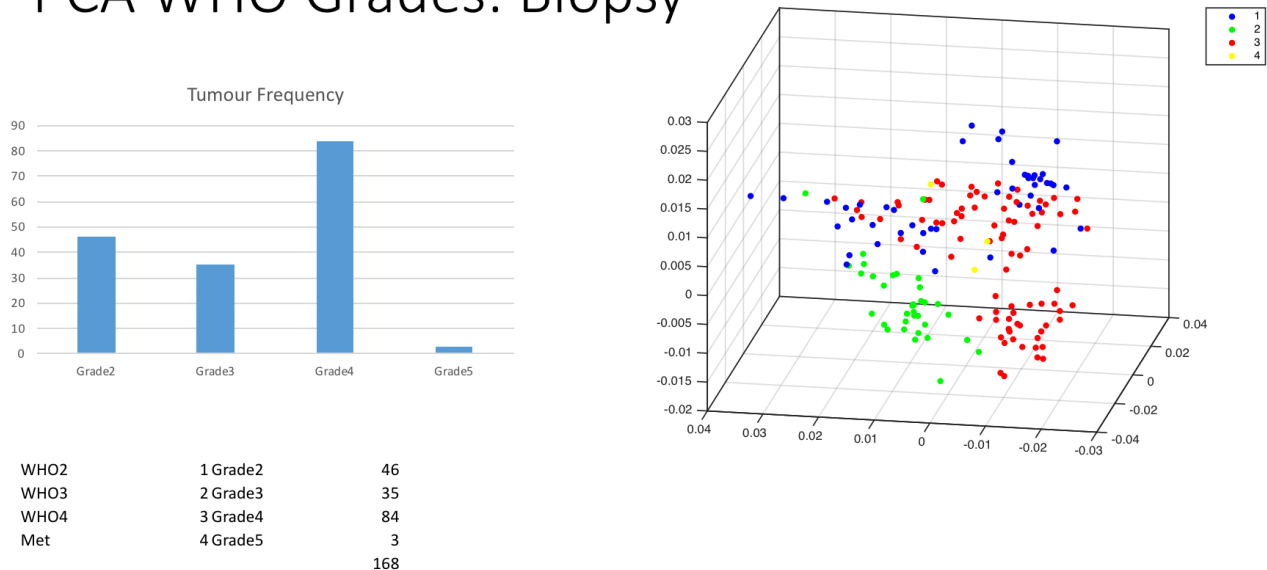


Figure 71. PCA plot of the first 3 principle components determining variation of WHO grades in biopsy-only spectra.

Excellent segregation is seen amongst the 4 WHO grades used to determine the clinic-pathological aggression and response to treatment of brain tumours.

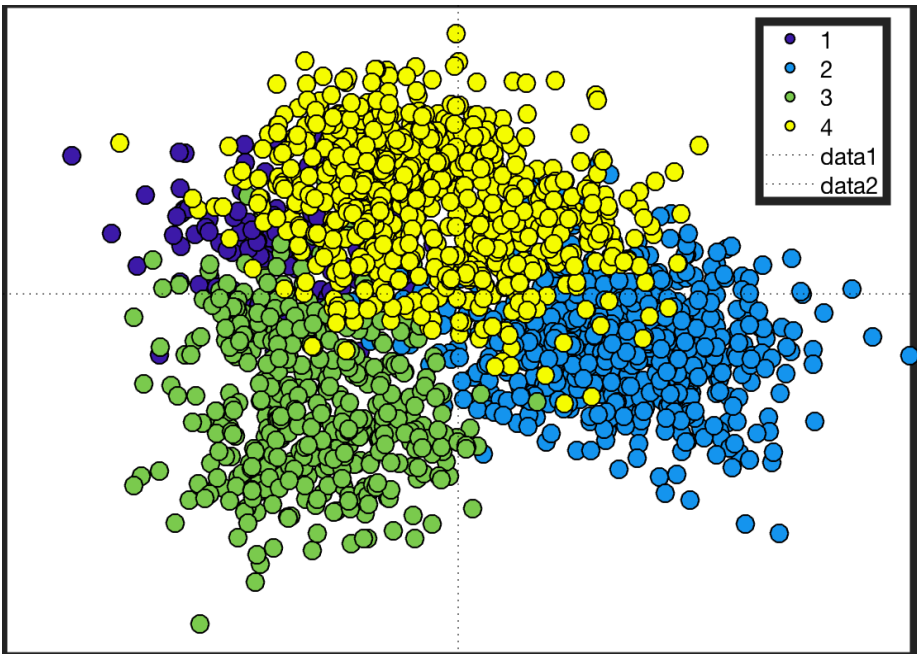


Figure 72. LDA plot of WHO grades in all Raman spectra collected in the study.
All 4 grades are represented here with excellent clustering within all grades.

The excellent segregation in both PCA and LDA suggests strong classification models are possible. PLS-DA and SVM models gave the highest accuracy. PLS with waveband selection predicted the pathology of these cases with a

sensitivity of 85% (95% CI of 82-88%), and a specificity of 80% (Table 14). See Table 15 for a breakdown of results for different tumour subclasses that I obtained.

Analyses Results		Test Data	G2, G3 and G4 Cases Only	
Training Data	Method	AUC (%)	Spec. (%)	Sens. (%)
Raman From G2, G3 or G4 Patients Only	Boosted Trees	90 (88 to 92)	69 (64 to 73)	85 (82 to 88)
	Partial Least Squares Full	89 (87 to 90)	78 (73 to 81)	85 (82 to 88)
	Partial Least Squares Select	92 (90 to 93)	84 (80 to 87)	85 (82 to 88)
Fluorescence From G2, G3 or G4 Patients Only	Boosted Trees	89 (87 to 91)	61 (56 to 65)	85 (82 to 88)
	Partial Least Squares Full	86 (84 to 88)	72 (67 to 76)	85 (82 to 88)
	Partial Least Squares Select	86 (84 to 88)	72 (67 to 76)	85 (82 to 88)
Raman and Fluorescence From G2, G3 or G4 Patients Only	Boosted Trees	94 (92 to 95)	87 (83 to 90)	85 (82 to 88)
	Partial Least Squares Full	94 (92 to 95)	89 (85 to 92)	85 (82 to 88)
	Partial Least Squares Select	95 (93 to 96)	92 (88 to 94)	85 (82 to 88)

Table 14. Classification of spectra acquired from patients with different tumour subclasses based on different multivariate methods.

Percentages in parenthesis are the 95% confidence intervals. Classification was based on training and test spectra drawn from a data subset of 953 normalized and combined Raman and Fluorescence spectra acquired from patients with WHO graded tumours (G2, G3 and G4). Two classifications were performed, a) using Raman spectra only and b) using combined Raman and Fluorescence spectra.

Results of the Statistical Analyses on 1171 Raman Spectra					
WHO Grade	Tumor Type	# Of Patients	Specificity (%)	Sensitivity (%)	Accuracy (%)
G2	G2 Astro	5	82	64	74
	G2 Oligo	2	73	90	82
	G2 Oligo/GBM	1	71	17	57
	G2 Type Totals	8	78	71	75
G3	G3	1	75	74	75
	G3 Oligo	1	NA	100	100
	G3 Oligo (GBM)	1	95	97	96
	G3 Oligoastro	1	89	NA	89
	G3 Type Totals	4	87	90	88
G4	GBM	8	76	94	89
G2,G3 and G4 Type Totals		20	80	85	82
Other	Hemangioma 4C	1	96	NA	96
	Lung Met	1	100	NA	100
	Meningioma	1	100	NA	100
	Radionecrosis	1	92	NA	92
	Choriocarcinoma	1	63	82	73
	Grade Totals	5	89	82	88
All Grades	All Types	25	82	85	84

Table 15. Classification of different tumour grade based on PLS analyses with waveband section.

Classification was based on training and test spectra drawn from the full data set of 1171 normalised Raman spectra acquired from patients with all tumour classes.

SVM produced superior results with an accuracy of 98% as follows:

Rattle Analysis using Raman and Fluorescence, SVM, 70/30 train/test split

Error matrix for the SVM model on Rattle All Data Normalized Ram and Flu Normalized for WHO grade analysis.csv [test] (counts):

	Predicted			
Actual	2	3	4	1
2	13	5	0	4
3	0	30	1	0
4	2	0	102	0
1	0	0	0	7

Error matrix for the SVM model on Rattle All Data Normalized Ram and Flu Normalized for WHO grade analysis.csv [test] (proportions):

	Predicted				
Actual	2	3	4	1	Error
2	0.48	0.00	0.01	0.00	0.03
3	0.00	0.11	0.00	0.00	0.03

4	0.01	0.00	0.36	0.00	0.02
1	0.00	0.00	0.00	0.02	0.00

Overall error: 2%, Averaged class error: 2%

Rattle timestamp: 2018-04-03 00:31:10 Babar Vaqas

=====

SVM performance seems robust as accuracy remains high at 96% when SVM is used for model building for the combined *in vivo* and *ex vivo* datasets:

Rattle, analysis of whole dataset (In vivo and Ex vivo combined) for WHO grade.

Error matrix for the SVM model on Rattle All Spectra Combined (in vivo and ex vivo) WHO Grade.csv [test] (counts):

		Predicted			
Actual	1	2	3	4	
1	40	0	1	6	
2	0	20	1	4	
3	1	0	13	2	4
4	3	2	0	20	8

Error matrix for the SVM model on Rattle All Spectra Combined (in vivo and ex vivo) WHO Grade.csv [test] (proportions):

		Predicted				
Actual	1	2	3	4	Error	
1	0.07	0.00	0.00	0.01	0.15	
2	0.00	0.33	0.00	0.01	0.02	
3	0.00	0.00	0.22	0.01	0.04	
4	0.00	0.00	0.00	0.34	0.02	

Overall error: 4%, Averaged class error: 6%

Rattle timestamp: 2018-04-03 23:20:28 Babar Vaqas

=====

The corresponding ROC has an AUC of 1.0 which suggests excellent classifier performance for WHO grade, as shown graphically in Figure 73.

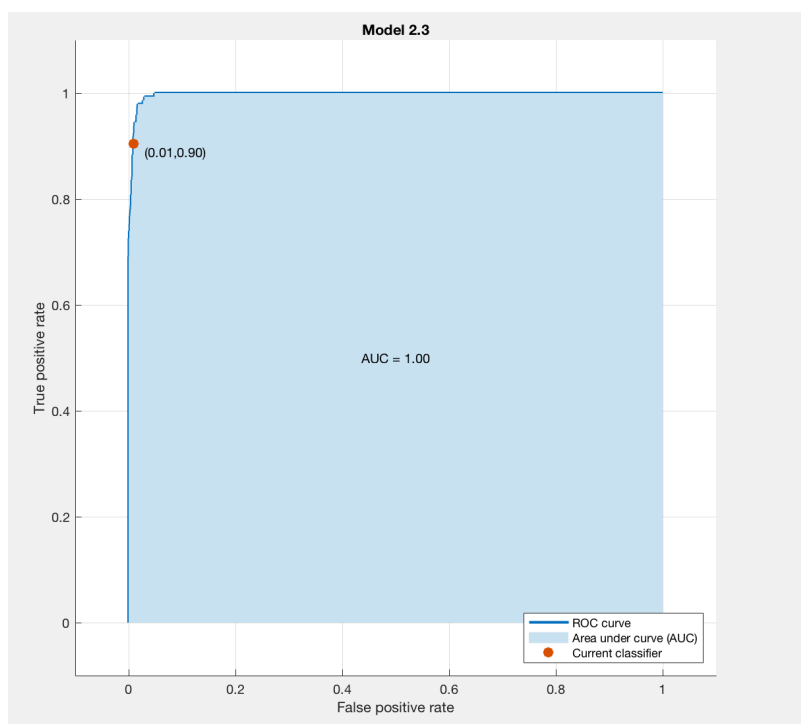


Figure 73. ROC curve of WHO grade as predicted by SVM models on *in vivo* and *ex vivo* data combined

The AUC is 1.0 suggesting an excellent classifier.

4.7.7.4. Summary

Tumour WHO grade seems to have a strong molecular basis in Raman spectroscopy with excellent classification based on my SVM models. Accuracy reached 96% with high specificity and sensitivity.

4.7.8. Tumour Type

4.7.8.1. Introduction

Moving on from simply detecting tumour tissue the next clinical question is whether Raman spectroscopy can be used to detect the type of tumour being operated on. There are 2 facets to this question. First, it will be determined if the can data be analysed to show a difference between glioma and non-glioma tissue as an important fundamental question. This relates to the ability of Raman spectroscopy to differentiate between intrinsic and extrinsic tumours of the brain. The second question is to what extent can specific disease types be characterised and accurately predicted from the Raman dataset.

The hypothesis is that these distinctions should be possible as these are tumour characteristics determined at the molecular level, and Raman spectroscopy obtains molecular information from tissue.

4.7.8.2. Average spectra

1863 glioma spectra and 574 non-glioma spectra were collected from a total of 2437 spectra. These spectra were pre-processed using intensity correction, normalisation with background fluorescence subtraction as described in section 3.6.2.1.

Average glioma and non-glioma spectra are shown in Figure 74.

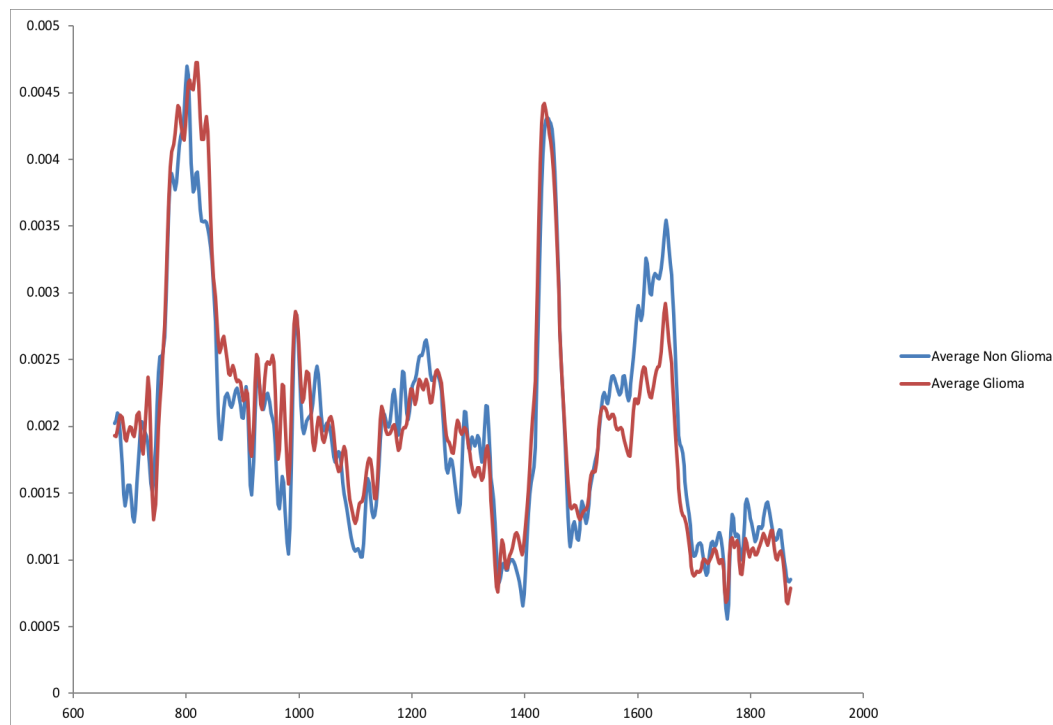


Figure 74. Averaged glioma spectra compared with non-glioma spectra from 29 patients.

Glioma spectrum average is shown in red and non-glioma spectrum average is in blue. Clear differences are observed between the 2 averaged spectra. In particular the lower wavenumber range 750-900 and the higher wavenumber range 1500-1650 are different between glioma and non-glioma tissue. The strong peak at 1450 is preserved.

The average differences I observed suggest that a classification model should be possible differentiating glioma from non-glioma.

A range of different tissue types were present amongst the 29 patients recruited in the Raman study. The classes studied are listed as follows:

Astrocytoma
Oligodendroglioma
Glioblastoma Multiforme
Lung Metastasis
Choriocarcinoma
Haemoangioblastoma
Radionecrosis

The average spectra for these tissue types are all unique and different from each other and are represented in Figure 69.

The high degree of tissue to tissue variation that exists suggests classification should be possible using Raman spectra, unless this variation is due primarily to noise.

4.7.8.3. Multivariate analysis

Unsupervised analysis of glioma versus non glioma spectra that I undertook shows no significant separation between the data, as shown graphically in Figure 75. *in vivo* and *ex vivo* data was combined for this analysis.

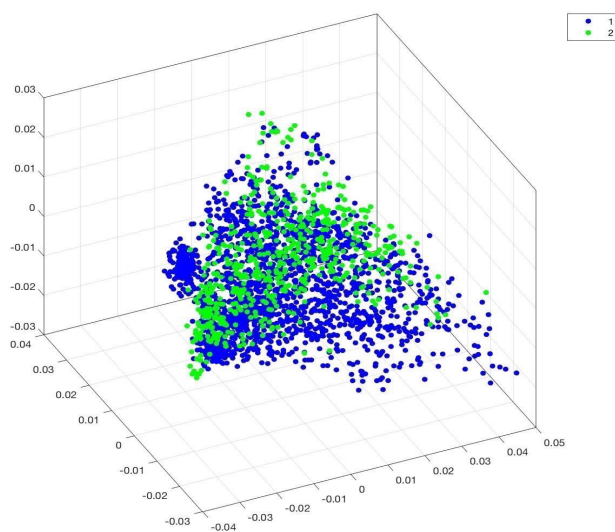


Figure 75. PCA of all *in vivo* and *ex vivo* spectra.
Glioma spectra are blue dots, non-glioma spectra and green dots.

Given there is no discernible separation on unsupervised analysis, I performed supervised LDA as shown in Figure 76. The results show good separation between the 2 classes.

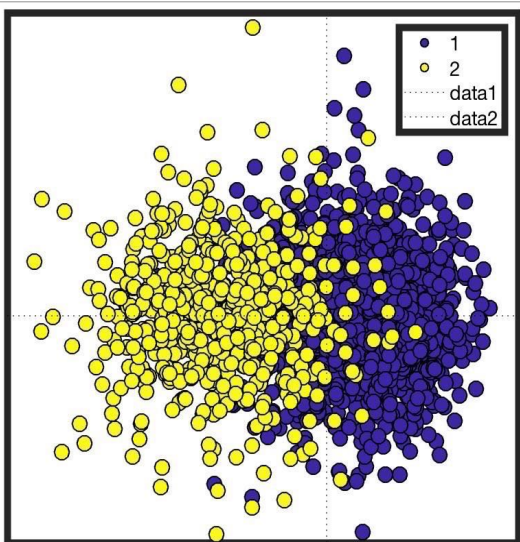


Figure 76. LDA plot of glioma vs non-glioma for all in vivo and ex vivo spectra combined.

Blue=glioma, yellow=non-glioma

An SVM model based on combined *in vivo* and *ex vivo* data shows excellent classification ability based on a large test set as shown below:

Rattle: SVM on in vivo and ex vivo data combined

This shows excellent classification using SVM.

Error matrix for the SVM model on Rattle All Spectra (in vivo and ex vivo) TvsN GvsNG.csv [test] (counts):

Actual	Predicted	
	Glioma	Non-Glioma
Glioma	369	11
Non-Glioma	11	97

Error matrix for the SVM model on Rattle All Spectra (in vivo and ex vivo) TvsN GvsNG.csv [test] (proportions):

Actual	Predicted		
	Glioma	Non-Glioma	Error
Glioma	0.76	0.02	0.03
Non-Glioma	0.02	0.20	0.10

Overall error: 5%, Averaged class error: 6%

Rattle timestamp: 2018-04-04 15:29:09 Babar Vaqas

=====

The next step in the analysis is to investigate differences in tumour histology diagnoses. I conducted unsupervised and supervised analyses.

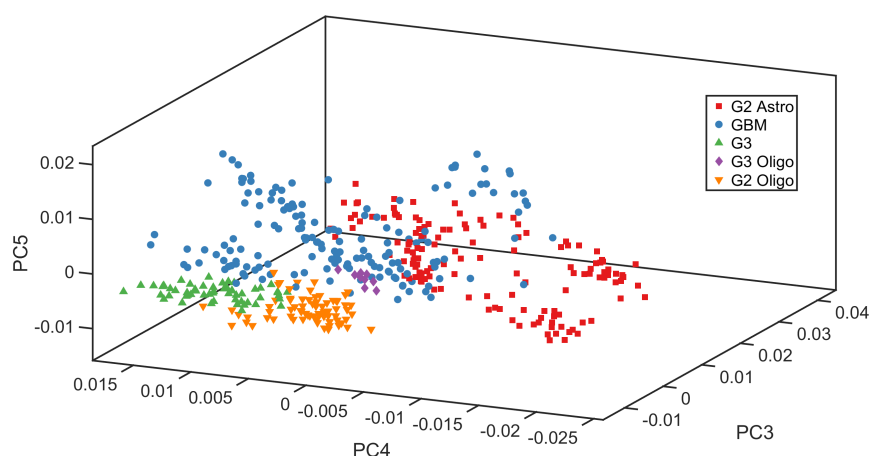


Figure 77. PCA plot showing PCA components 3,4 and 5.

Excellent segregation of spectra into tissue diagnostic groups is shown, with different grades of tumour within the same histological type segregating also.

Excellent segregation on PCA of different tumour types. Oligodendroglioma grades 2 and 3 also separate well, as does astrocytoma grade 2 and 3. PC 3,4,5 produce the best separation of diagnostic groups. This may be due to higher principle components explaining meaningless variation (e.g. due to blood) with diagnostic data contained in lower PC's. Notably, different histological grades of the same tumour segregate well also.

Not surprisingly given the excellent PCA clustering, LDA also shows excellent separation as shown in Figure 78.

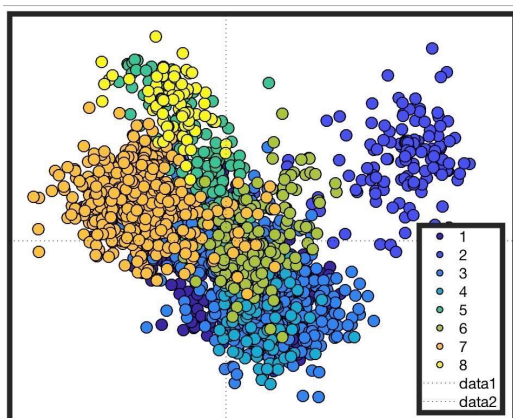


Figure 78. LDA plot of tumour type.

Excellent segregation of Raman spectra is seen based on tissue type. 8 diagnostic groups are represented as follows: 1 Astrocytoma, 2 Choriocarcinoma, 3 GBM, 4 Haemoangioblastoma, 5 Lung Met, 6 Meningioma, 7 Oligodendroglioma, 8 Radionecrosis.

Patients with WHO grade tumours (G2, G3 and G4) accounted for 953 out of the 1171 spectra. PLS with waveband selection predicted the pathology of these cases with a sensitivity of 85% (95% CI of 82-88%), and a specificity of 80% (Table 9). See Table 16 for a breakdown of results for different tumour subclasses.

Results of the Statistical Analyses on 1171 Raman Spectra					
WHO Grade	Tumor Type	# Of Patients	Specificity (%)	Sensitivity (%)	Accuracy (%)
G2	G2 Astro	5	82	64	74
	G2 Oligo	2	73	90	82
	G2 Oligo/GBM	1	71	17	57
	G2 Type Totals	8	78	71	75
G3	G3	1	75	74	75
	G3 Oligo	1	NA	100	100
	G3 Oligo (GBM)	1	95	97	96
	G3 Oligoastro	1	89	NA	89
	G3 Type Totals	4	87	90	88
G4	GBM	8	76	94	89
G2,G3 and G4 Type Totals		20	80	85	82
Other	Hemangioma 4C	1	96	NA	96
	Lung Met	1	100	NA	100
	Meningioma	1	100	NA	100
	Radionecrosis	1	92	NA	92
	Choriocarcinoma	1	63	82	73
	Grade Totals	5	89	82	88
All Grades	All Types	25	82	85	84

Table 16. Classification of different tumour grade based on PLS analyses with waveband section.

Classification was based on training and test spectra drawn from the full data set of 1171 normalized Raman spectra acquired from patients with all tumour classes.

SVM models also gave excellent classification performance with an accuracy of 94% as shown below:

Rattle: SVM model

80/20 data split for validation

All in vivo and ex vivo data combined

Error matrix for the SVM model on Rattle All Spectra (in vivo and ex vivo) TvsN GvsNG.csv [test] (proportions):

Actual	Predicted				
	Astrocytoma	Choriocarcinoma	GBM	Hemangioma	Lung
Met					
Astrocytoma	0.23	0.00	0.00	0.01	0.00
Choriocarcinoma	0.00	0.06	0.00	0.00	0.00
GBM	0.00	0.00	0.32	0.01	0.00
Hemangioma	0.00	0.00	0.00	0.02	0.00
Lung Met	0.00	0.00	0.00	0.00	0.04
Meningioma	0.00	0.00	0.01	0.00	0.00
Oligodendroglioma	0.00	0.00	0.01	0.00	0.00
Radionecrosis	0.00	0.00	0.00	0.00	0.00
Actual	Predicted				Error
	Meningioma	Oligodendroglioma	Radionecrosis		
Astrocytoma	0.00	0.0	0.00		0.06

Choriocarcinoma	0.00	0.0	0.00	0.00
GBM	0.00	0.0	0.00	0.02
Hemangioma	0.00	0.0	0.00	0.15
Lung Met	0.00	0.0	0.00	0.05
Meningioma	0.00	0.0	0.00	0.27
Oligodendroglioma	0.00	0.2	0.00	0.08
Radionecrosis	0.00	0.0	0.02	0.08

Overall error: 6%, Averaged class error: 9%

Rattle timestamp: 2018-04-04 17:01:22 Babar Vaqas

=====

4.7.8.4. Summary

Through analysis of Raman spectra using machine learning I have shown excellent ability to classify tissue on the basis of histological diagnosis, as hypothesised. The best performance is seen using SVM when combining *in vivo* and *ex vivo* data to reach an accuracy of 94% for a broad range of tissue diagnoses, including important differential diagnosis such as radionecrosis. These models are robust being built on over 2000 spectra and having high accuracy using a number of machine learning methods.

4.7.9. Tumour Margin

4.7.9.1. Introduction

So far, the Raman spectra I have analysed have shown a wide degree of variation between patients, with differences between spectra significant enough to classify them often best demonstrated using multivariate statistics and machine learning. The degree of within-patient variation that exists has important implications for surgical guidance during an operation and is therefore an important question to answer. A good model for answering this question is to consider the tumour vs normal distribution within a specific patient and to see if this is accurate when compared to histology (diagnostic accuracy) and surgical imaging before and at the end of surgery (margin detection). The findings of my analysis will be presented in this chapter.

4.7.9.2. Data analysis

Data from Patient 5 (P5) will be analysed as an example. The 'Normal Vs Tumour' model was built from the original *in vivo* Raman data but with P5 data excluded and then used to predict the identity of spectra collected during surgery in P5. The first step is to construct the model, this is then applied to P5 spectra and the performance of this model is then assessed using a confusion matrix.

All machine learning algorithms were tested with the best performing ones based on the confusion matrix being selected for model prediction on the 3D reconstructed tumour. SVM and BT produced the best models:

Data input: RattleTvsN.xls (dataset of all data, labelled Tumour and Normal)
 Model Builders: Rapidminer, Rattle
 Machine Learning Algorithm: Boosted Trees, SVM
 Confusion Matrix: SVM gave the best scoring matrix

```
PerformanceVector:
accuracy: 77.42%
ConfusionMatrix:
True:      Tumour      Normal
Tumour:    14          4
Normal:    3           10
precision: 77.78% (positive class: Tumour)
ConfusionMatrix:
True:      Normal      Tumour
Normal:    10          3
Tumour:    4           14
recall: 82.35% (positive class: Tumour)
ConfusionMatrix:
True:      Normal      Tumour
Normal:    10          3
Tumour:    4           14
AUC (optimistic): 0.727 (positive class: Tumour)
AUC: 0.727 (positive class: Tumour)
AUC (pessimistic): 0.727 (positive class: Tumour)
```

When the wrong predictions using the model are investigated it turns out most of the wrong predictions may be due to the model classifying tumour in normal MRI regions (at the brain/tumour interface) as shown in Figure 79.

Row No.	N vs T	predict...	confidence(Normal)	confidence(Tumour) ↓
5	Normal	Tumour	0.179	0.821
3	Normal	Tumour	0.219	0.781
4	Normal	Tumour	0.292	0.708
2	Normal	Tumour	0.305	0.695
7	Tumour	Normal	0.564	0.436
1	Tumour	Normal	0.658	0.342
6	Tumour	Normal	0.682	0.318

Figure 79. Predicted classes from SVM model for P5 data.

The model classifies one sample as tumour in a region classified as normal according to MRI.

Using boosted trees produces more wrong predictions

Row No.	N vs T	predict...	confidence(Normal)	confidence(Tumour)
1	Tumour	Normal	0.748	0.252
2	Normal	Tumour	0.116	0.884
3	Normal	Tumour	0.155	0.845
4	Normal	Tumour	0.181	0.819
5	Normal	Tumour	0.160	0.840
6	Normal	Tumour	0.192	0.808
7	Tumour	Normal	0.553	0.447
8	Tumour	Normal	0.834	0.166

Figure 80. Predicted classes from Ada Boost model for P5 data.

More wrong predictions result from using Boosted Trees.

However as is seen below, an important observation is that these wrong predictions are located at tumour margin and may represent areas which appear normal on MRI but which may actually contain tumour cells.

I projected this data onto 3D vector space overlying MRI data to understand the implication for detecting the tumour margin. I used 3D Slicer to create a 3D volume of the patient's tumour based on the preoperative volumetric MRI. Raman data points were identified and placed as ROI's (Regions of Interest) and then labels from the results of the predictive output of machine learning models were then displayed on top of this. This process can be performed in real time during surgery by connecting 3D Slicer to the navigation system through an Application Programming Interface (API) made available by the manufacturer. The outputs from the Raman model can then be labelled by the surgeon or an assistant. Figures 81 and Figure 82 show 2 outliers which were identified using this technique.

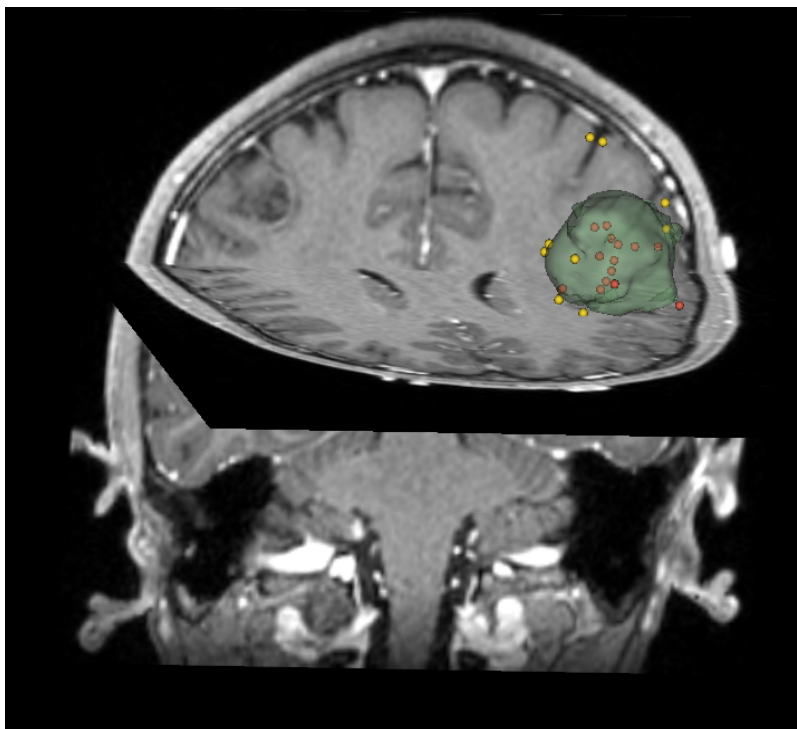


Figure 81. 3D visualisation of a brain tumour showing Raman data analysis points and highlighting the radiological tumour border, A/P view.

The radiological border is the edge of the tumour volume (green) whereas the chemical border is based on Raman sampling points, yellow=normal tissue,

red=tumour tissue. All red points lie inside the tumour volume apart from one outlier seen on the right side of the tumour.

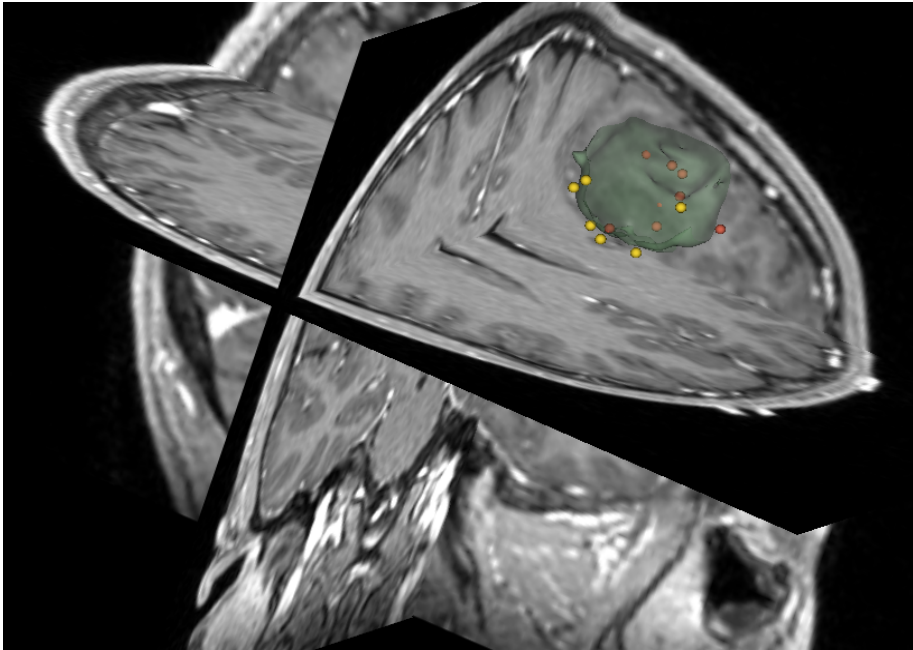


Figure 82. 3D visualisation of a brain tumour showing Raman data analysis points and highlighting the radiological tumour border, lateral view.

The radiological border is the edge of the tumour volume (green) whereas the chemical border is based on Raman sampling points, yellow=normal tissue, red=tumour tissue. All red points lie inside the tumour volume apart from one outlier seen on the right side of the tumour.

For patient number 5, the machine learning models used accurately identified tumour tissue from surrounding normal tissue. The false positive results may represent tumour tissue as these Raman points were not accompanied by formal histology as the readings were taken from eloquent brain tissue in the frontal lobe precluding a formal biopsy.

4.7.9.3. Summary

Analysis of predictive models of tumour versus normal tissue when used on single patient data show that the tumour and tumour border are delineated well in 3D space. I have shown that point Raman analysis correctly identified radiologically and histologically normal tissue surrounding the tumour, with 2 false positives which may represent tumour invasion into eloquent brain tissue.

4.7.10. Genomic Status

4.7.10.1. Introduction

Genomic alterations play a central role in gliomagenesis as discussed in section 2.2. Rapid genomic analysis would enable a precision medicine approach to glioma as discussed in section 2.4.3.

The main questions, therefore, to answer from data experiments were:

1. Is there a difference in Raman spectra from mutated vs non mutated tissue of different key driver mutations in glioma?
2. Can models be built to learn the difference from mutated vs non mutated tissue?
3. Can these models predict the mutational status of new samples of tissue with clinical accuracy?

The hypothesis is that since cancer is defined as a genetic condition, the principle underlying cause of variation between tissue samples will be molecular. Given that Raman spectroscopy is a molecular analysis technique, spectra from mutated tissue will therefore be unique and specific to the mutations present.

Genetic information combined with *in vivo* spectra and tissue biopsy samples were available for 17 patients in this study. Their demographic and histological diagnoses are summarised in Table 17.

WHO Grade	Diagnosis	Number of Patients
2		6
	Astrocytoma	5
	Oligodendroglioma	1
3		2
	Astrocytoma	1
	Oligodendroglioma	1
4		9
	GBM	9
Age: median (range)	42 (22-79)	
	Total	17

Table 17. Patient demographics and histological diagnosis for genetic analysis.

17 patients were enrolled in this study with a range of low and high grade tumours. All tissue samples were subjected to molecular histopathological analysis in line with the World Health Organisation 2016 classification of brain tumours. GBM=Glioblastoma Multiforme.

A total of 471 spectra were collected matched to *in vivo* biopsy tissue. This tissue was tested for one or more of the following genetic changes: IDH mutation, ATRX expression and MGMT methylation status. The number of patients in each group and

number of spectra are summarised in Table 18.

Molecular Test	Status	Number of Patients	Number of Spectra
IDH	Mutated	10	258
	Wild Type	7	171
ATRX	Loss	7	164
	Retained	9	230
MGMT	Methylated	7	128
	Unmethylated	10	186
Total		17	471

Table 18. Summary of IDH, ATRX and MGMT mutational states and the number of spectra collected from tissue for molecular analysis.

A total of 471 spectra were collected from 17 patients. IDH= Isocitrate Dehydrogenase gene, ATRX=Alpha Thalassemia/Mental Retardation Syndrome X-linked gene, MGMT=O6-methylguanine DNA methyltransferase.

4.7.10.2. IDH

Analysis of average spectra from tissue with IDH Wildtype (WT) and Mutant (MT) status reveals significant differences. Figure 83 shows the average spectra together with the difference spectrum with statistically significant differences highlighted in red.

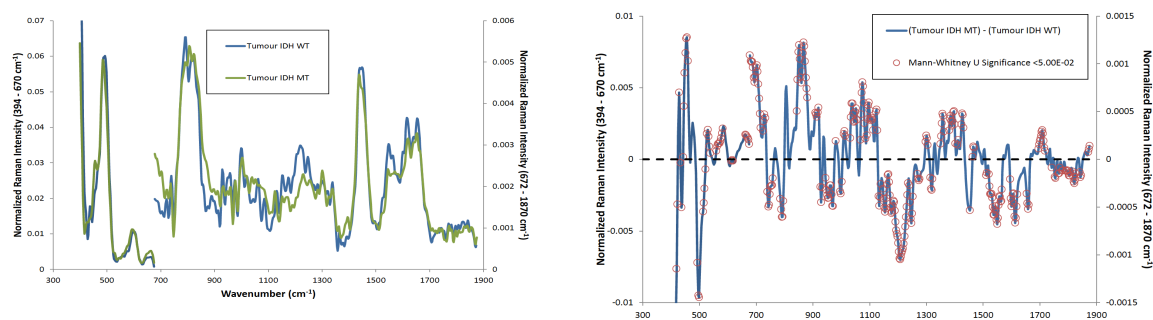


Figure 83. Average spectra of IDH mutated and non mutated tissue (left) and the difference spectrum (right).

Statistically significant differences between the spectra are shown in red.

It can be seen that significant differences exist along the whole range of wavenumbers. The next step is to identify a reference spectra from the average difference spectrum if possible. This would be suggestive of the presence of the reference molecule for further investigation.

I performed supervised hierarchical cluster analysis to reduce variability in the spectra caused by tumour gene mutation and grade as shown in Figure 84. This improved the sensitivity and specificity further (to 100%) suggesting stronger

differences between mutant and non mutant spectra in these groups.

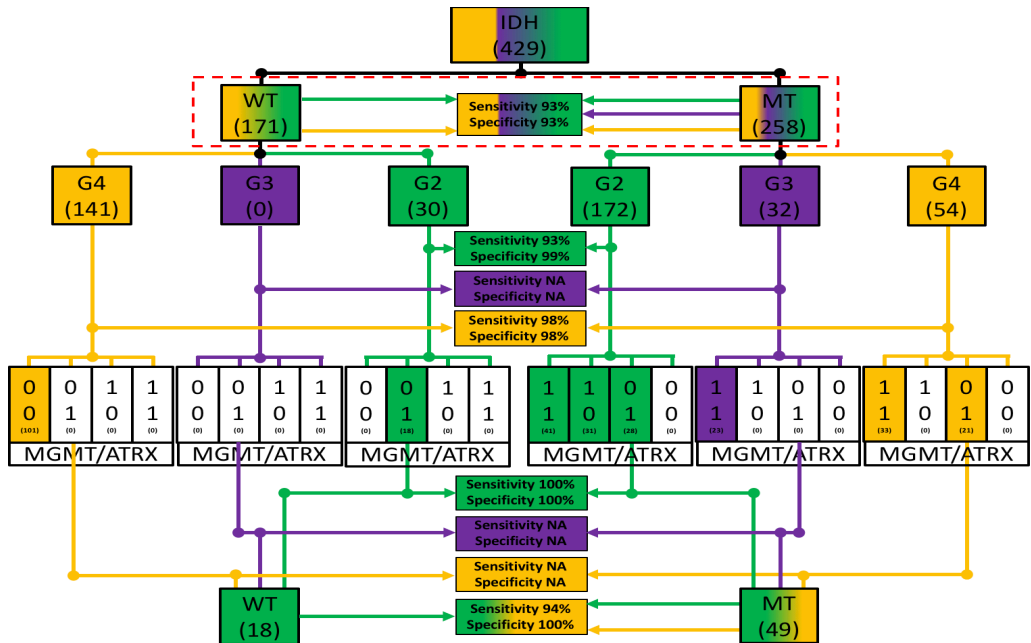


Figure 84. IDH spectral comparisons and associated sensitivity and specificities for spectra grouped by grade.

Supervised hierarchical cluster analysis was performed to increase sensitivity and specificity. The best performing comparisons are in green. Numbers in brackets represent numbers of spectra.

Using the average spectrum from successive hierarchical groups which reduce spectral variability produces a progressively closer match of the number of peaks with the reference 2-HG spectrum, as shown in Figure 85. The spectra become progressively less variable.

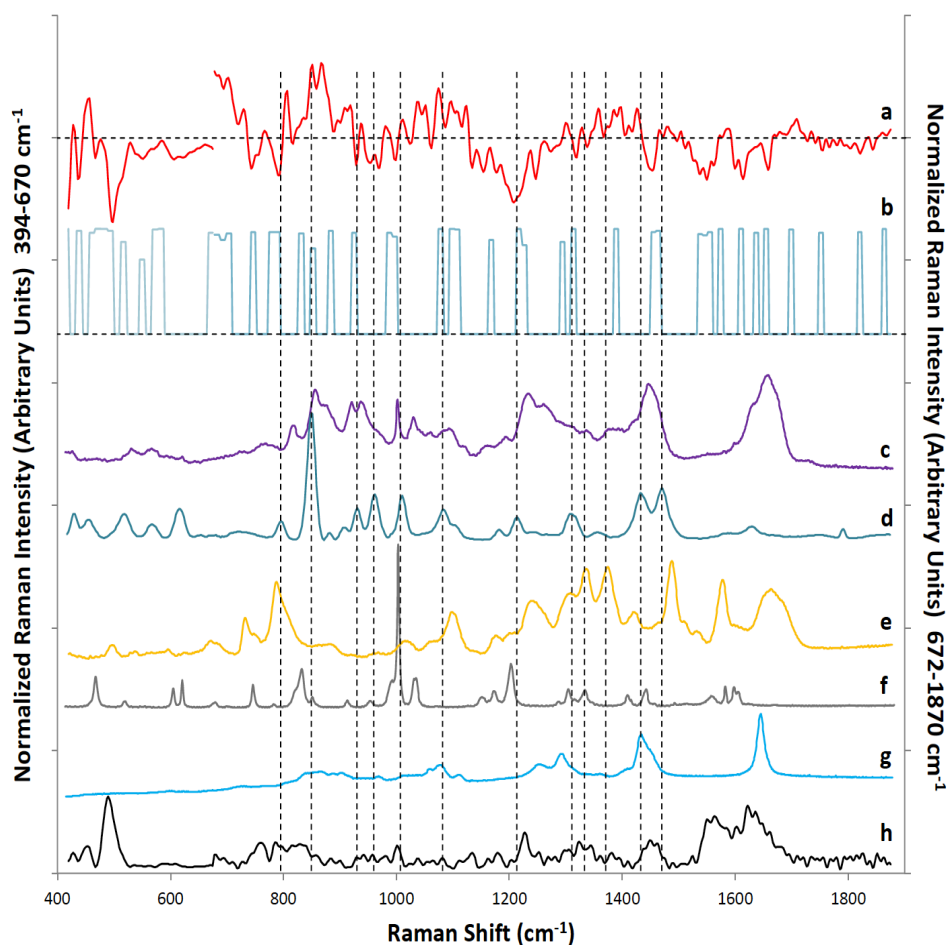


Figure 85. Comparison of successive difference spectra for IDH mutation with reference spectra.

a) all 429 tumour spectra tested for IDH mutation; b) all 202 G2 tumour spectra tested for IDH mutation; c) all 195 G4 tumour spectra tested for IDH mutation; d) all 46 G2 tumour spectra tested for IDH mutation and with ATRX=1 and MGMT=0; e) all 67 G2 and G4 tumour spectra tested for IDH mutation and with ATRX=1 and MGMT=0. Reference spectra: f) collagen; g) 2-HG; h) DNA

Comparing the IDH difference spectrum from all Grade 2 spectra with the 2-HG reference spectrum measured using the *in vivo* Raman system shows that 15 out of 19 peaks match (Figure 86).

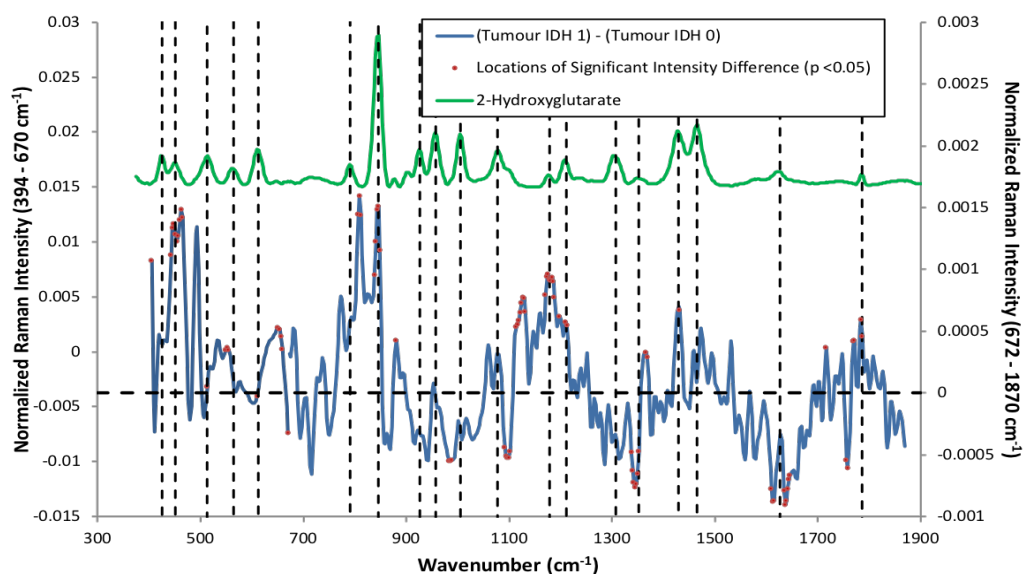


Figure 86. The effect of the IDH MT is investigated in spectra from all tumours but again only those with ATRX 1, and MGMT 0.

There is a slightly better correlation between the 2HG reference spectrum and the peaks in the difference spectra (15 out of 19 main peaks), the difference is better but still statistically rather weak. Possible explanation: 2HG increases in MT tissue are still small compared with the inter-patient variation because only 3 patients (P5, WT and P8, P14, MT) meet the required restrictions of ATRX 1, and MGMT 0.

Using the second order derivative of the IDH difference spectrum used in Figure 86 and 2-HG, all peaks match (green) apart from 3 red peaks as shown in Figure 87.

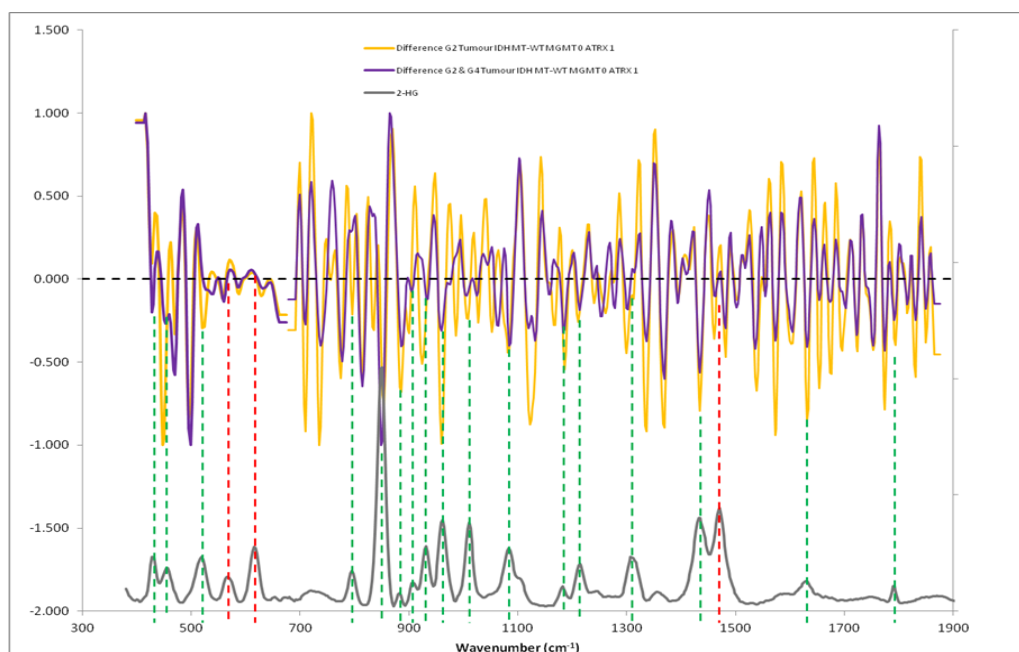


Figure 87. Second order derivative peaks and comparison with 2-HG reference spectrum.

All apart from 3 peaks match producing an excellent fit between the IDH second order derivative spectrum and 2-HG reference spectrum.

These findings suggest, but do not prove, that 2-HG may be responsible for the difference being observed between IDH MT and WT spectra. In order to investigate the dataset further and build predictive models, dimensionality reduction using PCA was first performed.

PCA analysis shows excellent separation between IDH WT and MT spectra using the first 3 principle components as shown in Figure 88.

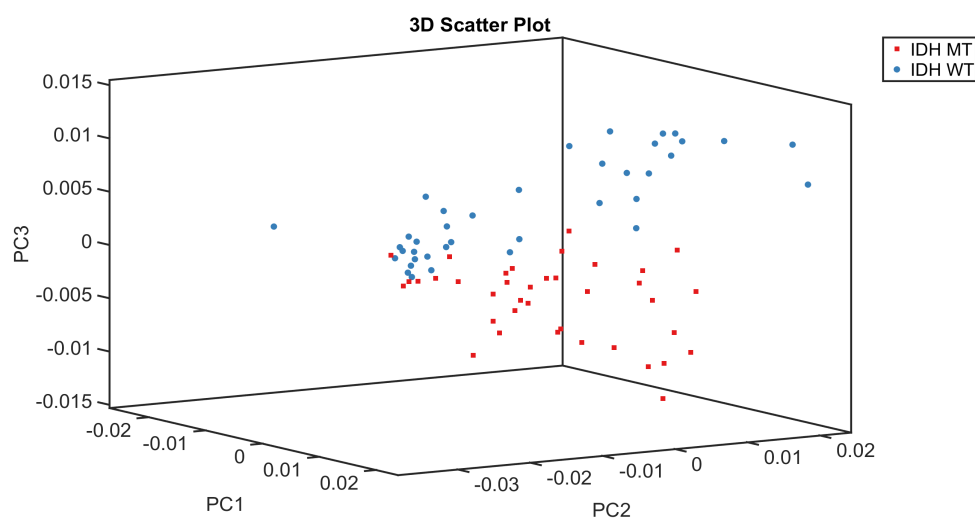


Figure 88. PCA plot of IDH MT vs IDH WT.

There is excellent segregation between IDH mutated and wild type spectra with hardly any overlap. IDH MT=red squares, IDH WT= blue circles.

The loadings plot in Figure 89 suggests DNA, protein and lipid regions of the Raman spectrum are responsible for the differences seen.

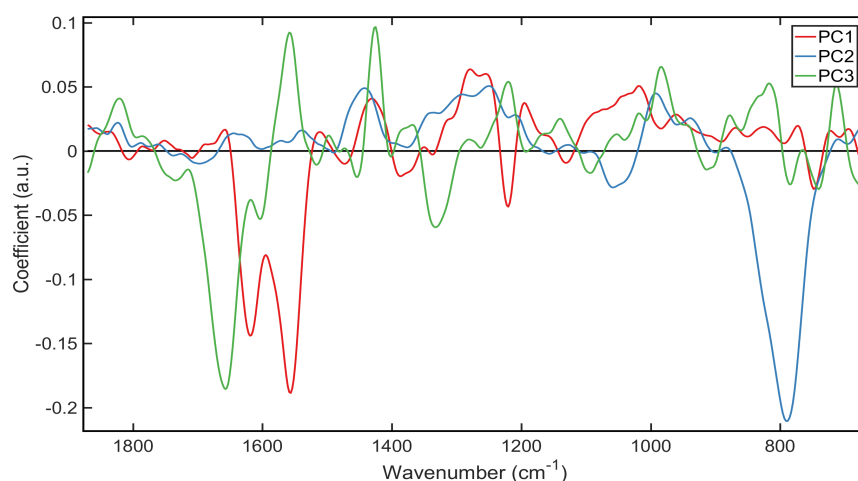


Figure 89. Loadings plot of the PCA analysis in Figure 88.

This shows the most significant wavenumbers accounting for the difference between IDH mutant and IDH wild type spectra. These wavenumbers span the DNA, protein and lipid regions of the Raman spectrum.

Putative identities of the peaks seen in some of the key principle components are shown in Table 19.

IDH	PCA 3 (cm ⁻¹)	PCA 4 (cm ⁻¹)	PCA 18 (cm ⁻¹)
Peak 1	1000 (phenylalanine)	1300 (Lipid)	609 (Protein)
Peak 2	1247 (Amide III)	702 (Phospholipids/cholesterol)	675 (Cholesterol/protein)
Peak 3	848 (DNA/RNA)	1008 (phenylalanine)	534 (Cholesterol/protein)

Table 19. Key peaks seen in principle components 3,4 and 18.

The key lipid, protein and DNA/RNA peaks seen are listed. Peak identity is based on already published literature summarised in Figure 160.

The fact that the data separates on variance between principle components suggests that strong classifiers should be possible. I used ensemble machine learning methods to produce the most accurate predictive models, in particular the Ada Boost and Random Forest models, produced excellent models based on cross validation as shown below:

Error matrix for the Ada Boost model on IDH Tumour Rattle.csv [test] (counts):

	Predicted	
Actual	0	1
0	36	2
1	2	71

Error matrix for the Ada Boost model on IDH Tumour Rattle.csv [test] (proportions):

	Predicted		
Actual	0	1	Error
0	0.32	0.02	0.05
1	0.02	0.64	0.03

Overall error: 4%, Averaged class error: 4%

Rattle timestamp: 2018-04-14 23:07:02 Babar Vaqas

=====

Error matrix for the Random Forest model on IDH Tumour Rattle.csv [test] (counts):

	Predicted	
Actual	0	1
0	36	2
1	0	73

Error matrix for the Random Forest model on IDH Tumour Rattle.csv [test]

(proportions):

Actual	Predicted		Error
	0	1	
0	0.32	0.02	0.05
1	0.00	0.66	0.00

Overall error: 2%, Averaged class error: 2%

Using SVM supervised learning a model was created using Rattle in R. The ROC curve has an AUC of 1 showing this model to have excellent classification accuracy. SVM confusion matrix:

=====

Error matrix for the SVM model on IDH Tumour Rattle.csv [test] (counts):

Actual	Predicted	
	0	1
0	37	1
1	2	71

Error matrix for the SVM model on IDH Tumour Rattle.csv [test] (proportions):

Actual	Predicted		Error
	0	1	
0	0.33	0.01	0.03
1	0.02	0.64	0.03

Overall error: 3%, Averaged class error: 3%

Rattle timestamp: 2018-04-14 23:07:04 Babar Vaqas

=====

=====

All ensemble methods and SVM are compared in Figure 90 demonstrating the superiority of SVM, Ada and RF. Precision is the same as positive predictive value, and recall is the sensitivity of the model.

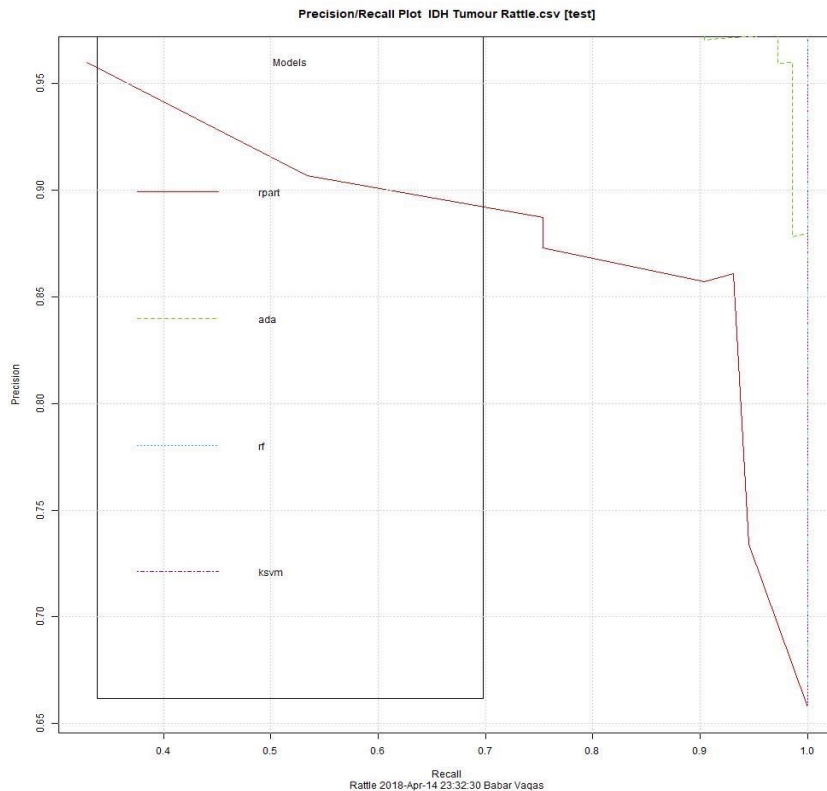


Figure 90. Precision/Recall plot comparing SVM and ensemble methods model performance.

Superior performance of SVM and RF is shown (labelled as ksvm and rf respectively on the chart).

PLS-DA modelling is also a very accurate modelling system for predicting IDH status. Calculating posterior probabilities using this model shows excellent ability to discriminate based on IDH mutational status irrespective of tumour grade as shown in Figure 91. Better performance is however seen for lower grade tumours (WHO grades 2 and 3).

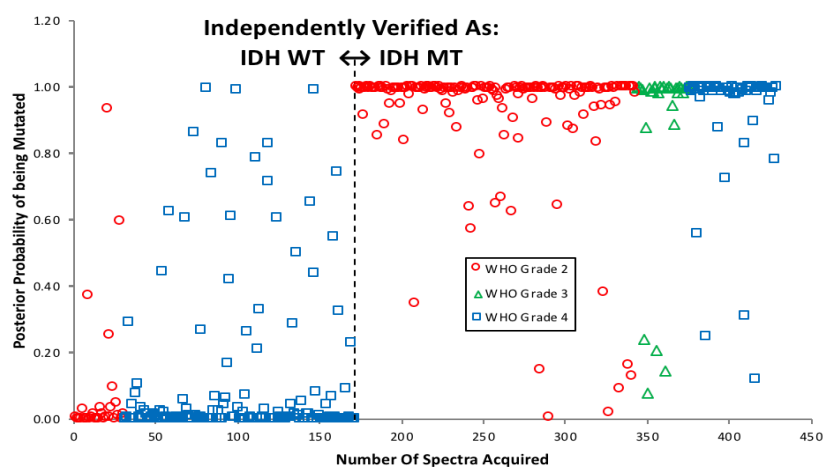


Figure 91. Probability of IDH mutation in brain tissue using PCA-LDA prediction models.

Good separation between IDH WT and MT tissues is seen with best separation for WHO Grades 2 and 3.

The frequency distributions of IDH WT and MT populations show clear differences as shown in Figures 92 and Figure 93 suggesting a strong ability of PCA-LDA to distinguish IDH MT from WT spectra.

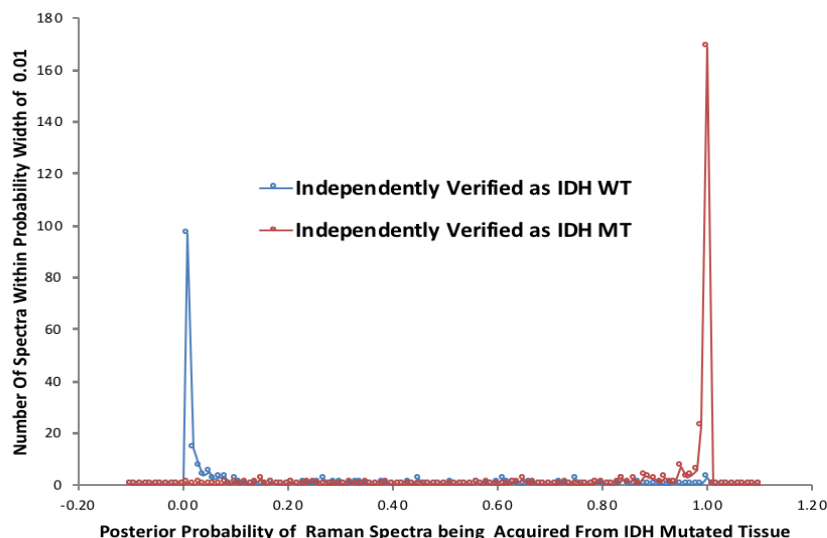


Figure 92. Frequency distribution in the probability of brain tissue having the IDH mutation using PCA-LDA.

A clear separation between IDH WT (blue) and MT (red) as identified by histopathology is seen, with little overlap.

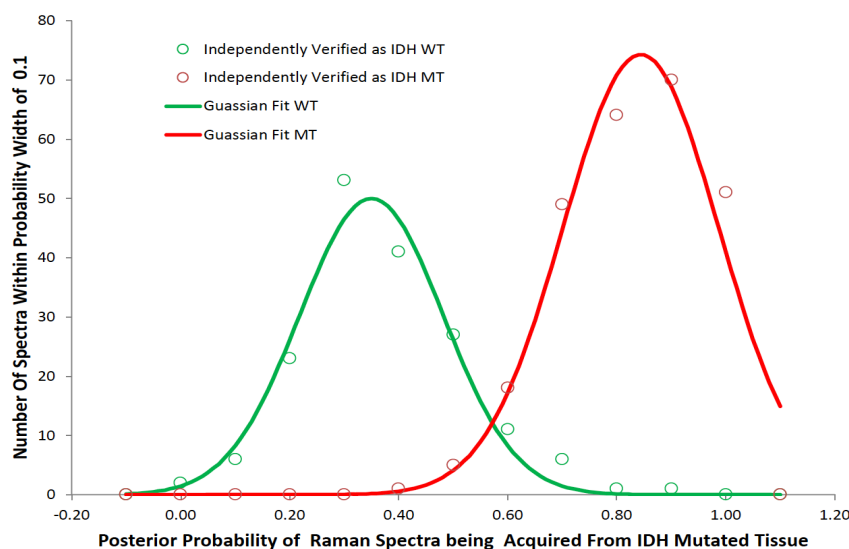


Figure 93. Distribution of number of spectra within probability width of 0.1 against posterior probability of spectra acquired from IDH mutated tissue.

Clear separation of IDH MT and WT is seen with little overlap between the 2 populations of spectra identified.

Classification models based on PCA-LDA were robust, as expected from the exploratory data analysis above. Figure 94 shows the ROC and AUC for this

classifier. Sensitivity and specificity was 93% with an AUC of 98%.

ROC Distribution In The Probability of Brain Tissue Having The IDH Mutation Using PCA-LDA Analyses of Raman Spectra From the Tissue

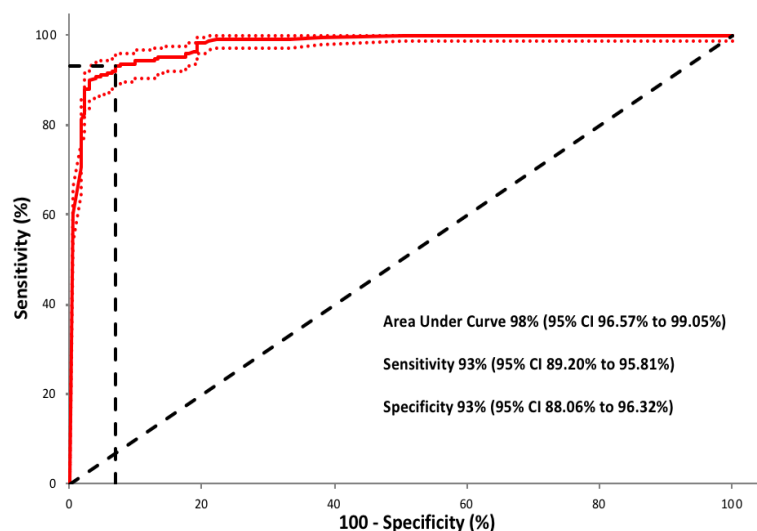


Figure 94. ROC curve and AUC for the PLS-LDA model when predicting IDH mutation.

With an AUC of 98% and a sensitivity and specificity of 93% the PCA-LDA model shows excellent ability to predict the IDH status of tissue being analysed.

4.7.10.3. ATRX

Analysis of the average spectra for ATRX mutation show obvious visual differences between mutated and non mutated spectra as shown in Figure 95. The difference spectrum shown in Figure 96 shows that there are multiple areas of statistical significant difference between the 2 spectra.

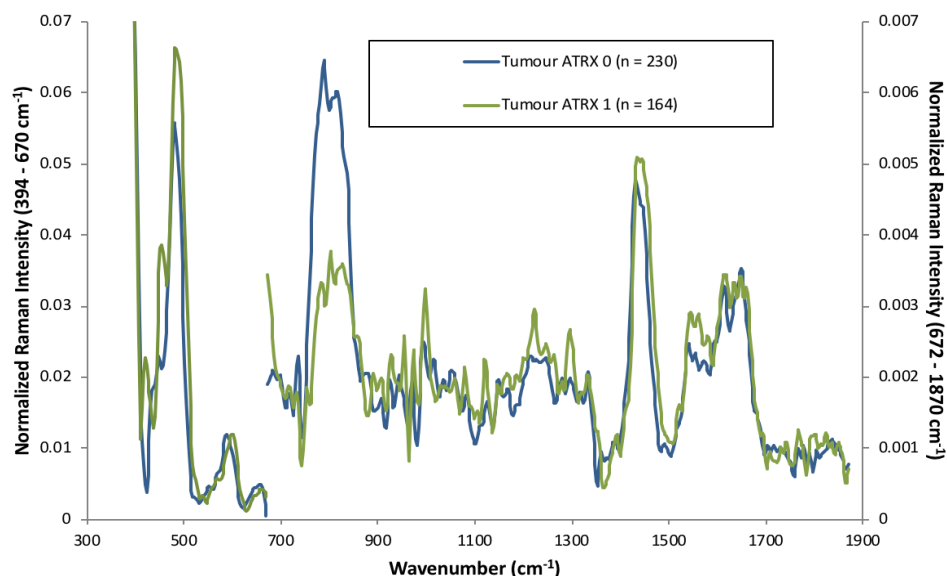


Figure 95. Average spectra from ATRX WT and ATRX MT tissue.

Spectra from ATRX mutated (green) and ATRX non-mutated (blue) tissue are shown superimposed in the 390-1800 wavenumber range. The break in spectra at 690 is due to the optical filters used in the system. Differences are visible throughout the DNA and protein regions of the Raman spectrum.

The difference spectrum matches closely with the DNA reference spectrum suggesting the differences may be due to DNA. Nearly all of the DNA peaks match changes in the difference spectrum for ATRX mutation.

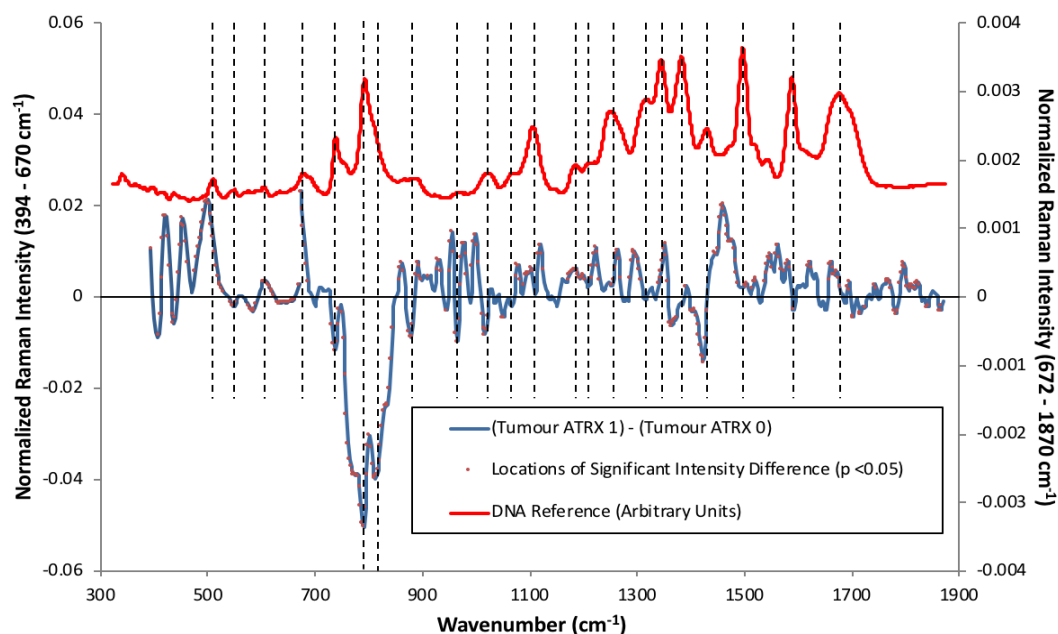


Figure 96. ATRX Raman difference spectrum compared to DNA reference spectrum.

Nearly all DNA peaks match the ATRX difference spectrum peaks.

When the ATRX spectra are stratified for comparison according to WHO grade as shown in Figure 97, the classifier performance is improved, and the resulting difference spectra consistently correlate with the DNA reference spectrum suggesting that this may be a true correlation as shown in Figure 98.

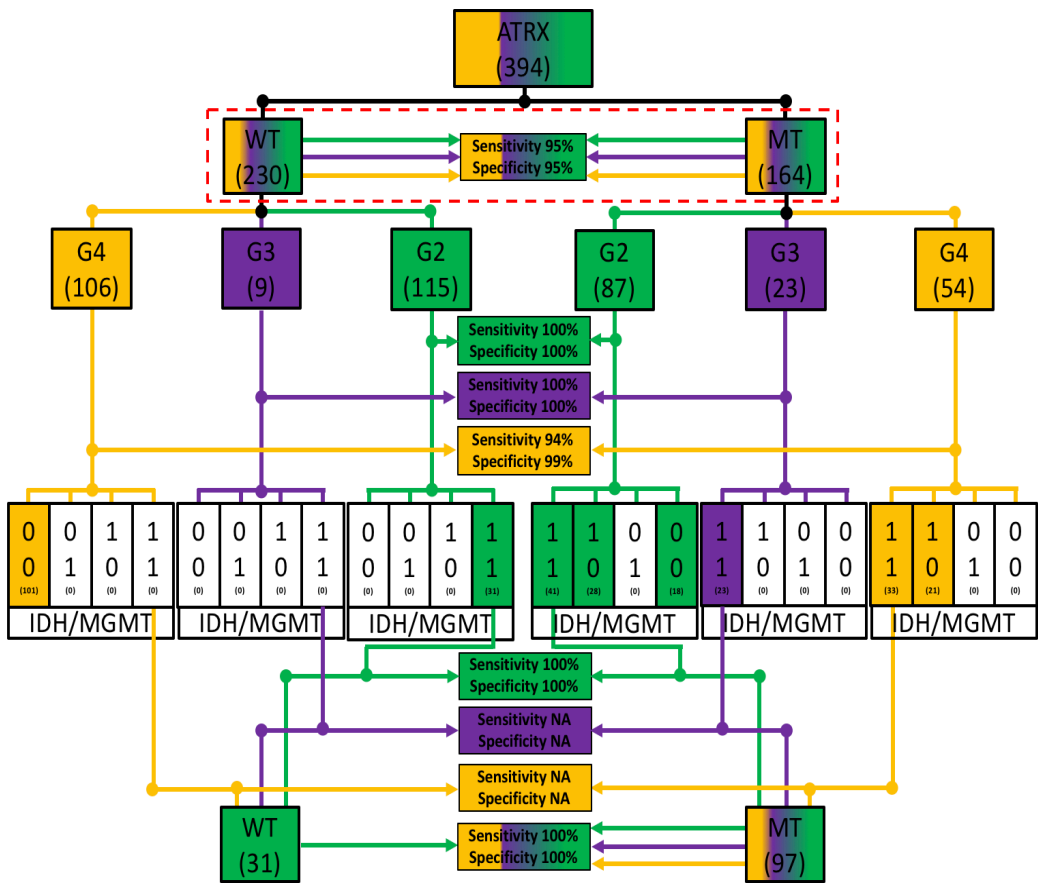


Figure 97. Supervised hierarchical cluster analysis of ATRX spectra. When stratifying spectra according to tumour grade and the presence of other mutations improved classifier performance is seen.

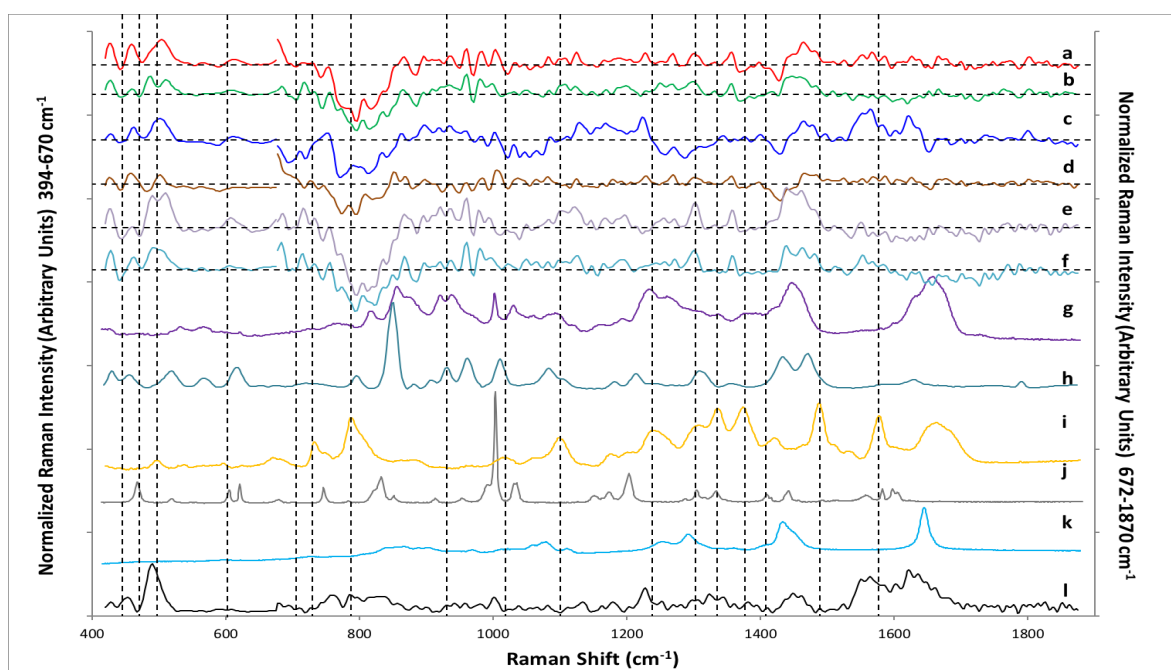


Figure 98. Comparison of successive difference spectra for ATRX mutation with reference spectra.

a) all 394 tumour spectra tested for ATRX mutation; b) all 202 G2 tumour spectra tested for ATRX mutation; c) all 32 G3 tumour spectra tested for ATRX mutation; d) all 160 G4 tumour spectra tested for ATRX mutation; e) all 72 G2 tumour spectra tested for ATRX mutation and with MGMT=1 and IDH=1. f) all 128 G2, G3 and G4 tumour spectra tested for ATRX mutation and with MGMT=1 and IDH=1. g) collagen; h) 2-HG; i) DNA; j) phenylalanine; k) Lipid (oleic acid); l) human blood. In this case there appears to be a significant correlation in the difference spectra with DNA peaks. And although the spectra do change as the number of variables that go into making these difference spectra (tumour type, ATRX and IDH status) are reduced the correlation with DNA seems to be consistent (vertical dashed lines).

Unsupervised analysis of the ATRX data using PCA shows some separation as shown in Figure 99, although this is not as good as with IDH mutation.

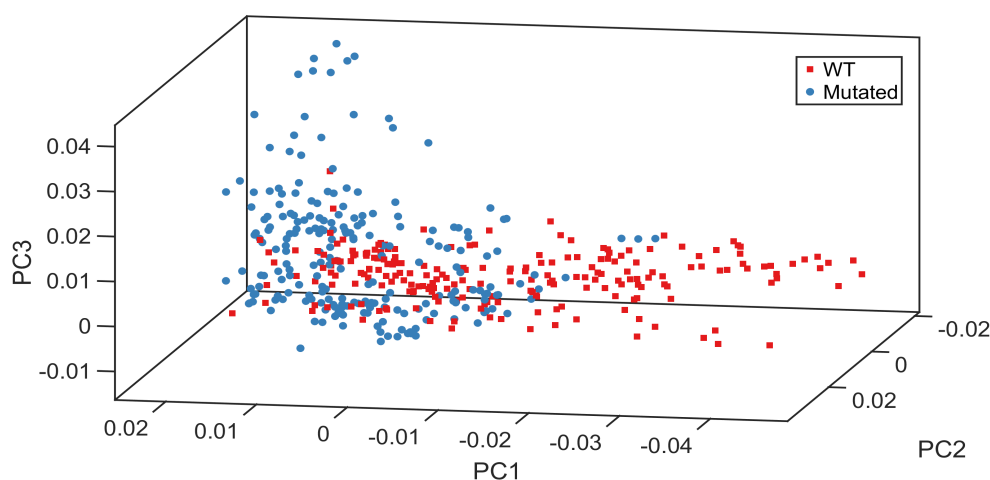


Figure 99. PCA plot of ATRX MT vs ATRX WT.

There is good segregation between ATRX mutated and wild type spectra with hardly any overlap. ATRX WT=red squares, ATRX Mutated= blue circles.

PCA loadings for PC 3,4 and 5 (avoiding PC1,2 and 3 due to blood signal) in Figure 100 show that the wavenumbers most responsible for the variation between mutated and non-mutated ATRX tissue spectra span the DNA, lipid and protein domains.

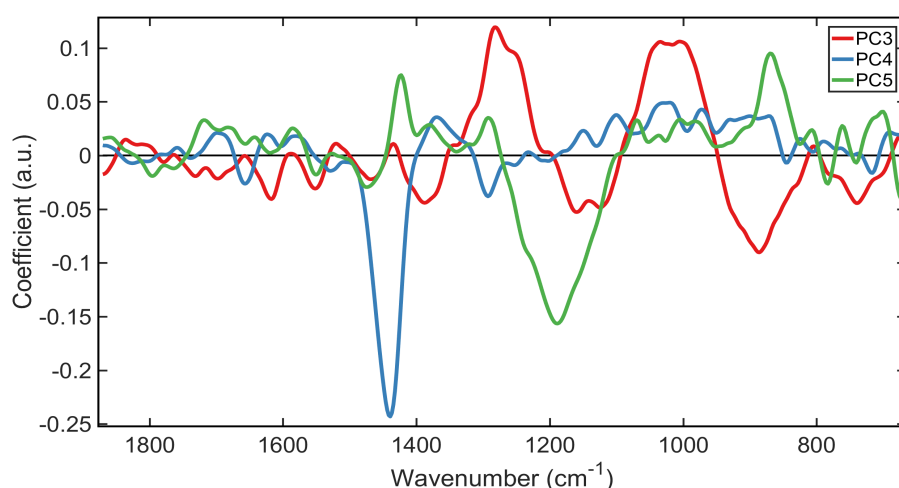


Figure 100. Loadings plot of the PCA analysis for ATRX mutation.

This shows the most significant wavenumbers accounting for the difference between IDH mutant and IDH wild type spectra. These wavenumbers span the DNA, protein and lipid regions of the Raman spectrum.

The peak identifiers responsible for the most variation between the classes are shown in Table 20.

ATRX	PCA 2	PCA 5	PCA 7
Peak 1	840 (Nucleic Acid)	1445 (Amide III)	1740 (Lipid)
Peak 2	775 (Nucleic Acid)	635 (Protein)	1718 (Lipid)
Peak 3	512 (Protein)	498 (Glycogen)	427 (Cholesterol)

Table 20. Key peaks seen in principle components 2,5 and 7.

The key lipid, protein and DNA/RNA peaks seen are listed. Peak identity is based on already published literature summarised in Figure 160.

The best predictive models from this data were RF, SVM, PCA=LDA and PLS-DA. The RF and SVM results are shown below:

Error matrix for the Random Forest model on ATRX Rattle.csv [test] (counts):

	Predicted
Actual 0	1

```

0 52 3
1 7 72

```

Error matrix for the Random Forest model on ATRX Rattle.csv [test] (proportions):

```

      Predicted
Actual  0   1 Error
0 0.39 0.02 0.05
1 0.05 0.54 0.09

```

Overall error: 7%, Averaged class error: 7%

Rattle timestamp: 2018-04-15 01:21:55 Babar Vaqas

```

=====
=====

```

Error matrix for the SVM model on ATRX Rattle.csv [test] (counts):

```

      Predicted
Actual 0   1
0 55 0
1 1 78

```

Error matrix for the SVM model on ATRX Rattle.csv [test] (proportions):

```

      Predicted
Actual  0   1 Error
0 0.41 0.00 0.00
1 0.01 0.58 0.01

```

Overall error: 1%, Averaged class error: 0%

Rattle timestamp: 2018-04-15 01:21:56 Babar Vaqas

```

=====
=====

```

Both models performed well when compared to BT and Recursive Partitioning as shown in Figure 101 reaching close to 100% sensitivity and specificity.

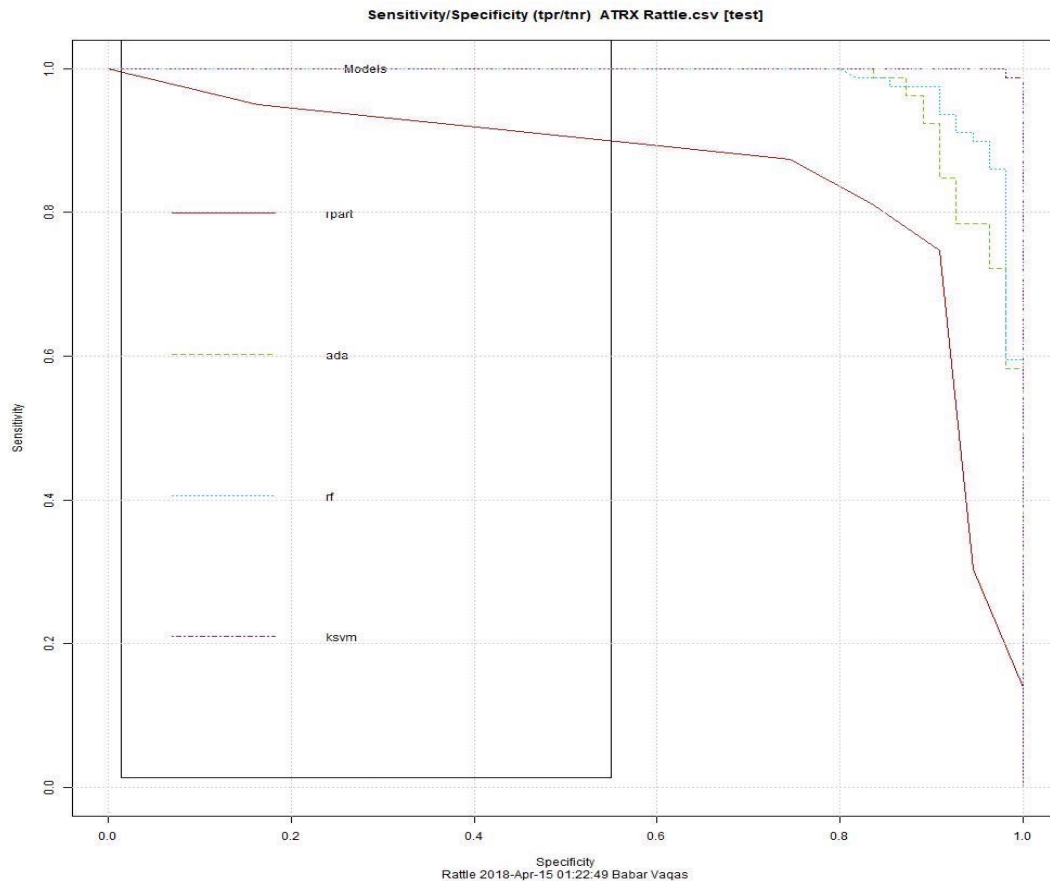


Figure 101. Sensitivity/Specificity plot comparing SVM and ensemble methods model performance for ATRX status.

Superior performance of SVM and RF is shown (labelled as ksvm, red dotted line, and rf, blue dotted line, respectively on the chart).

PCA-LDA methods also produced excellent discrimination ability between the 2 classes as shown in the posterior probability plot in Figure 102. There is excellent separation between the 2 classes with minimal cross over. WHO grade 2 and 3 showed the best separation.

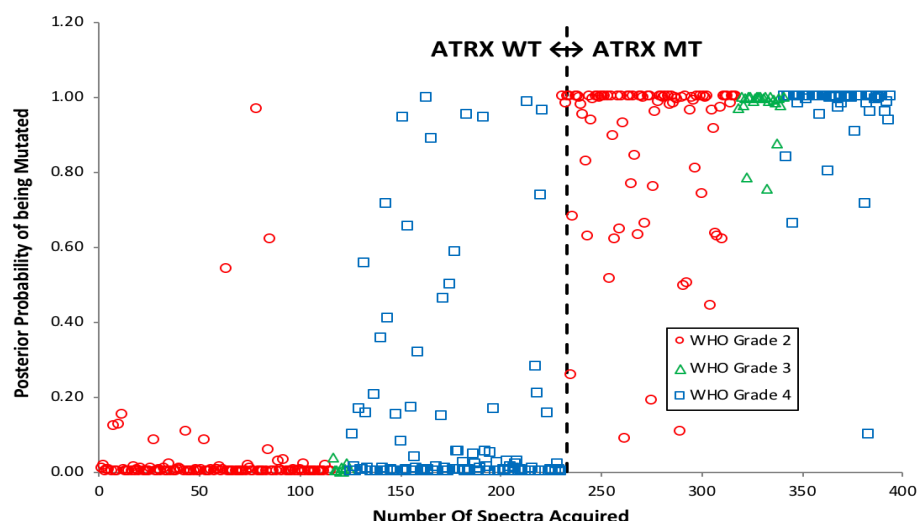


Figure 102. Probability of ATRX mutation in brain tissue using PCA-LDA prediction models.

Good separation between IDH WT and MT tissues is seen with best separation for WHO Grades 2 and 3.

Furthermore, when looking at the frequency distribution of the probabilities in the PCA-LDA model the level of certainty for the classification is high as shown in Figure 103.

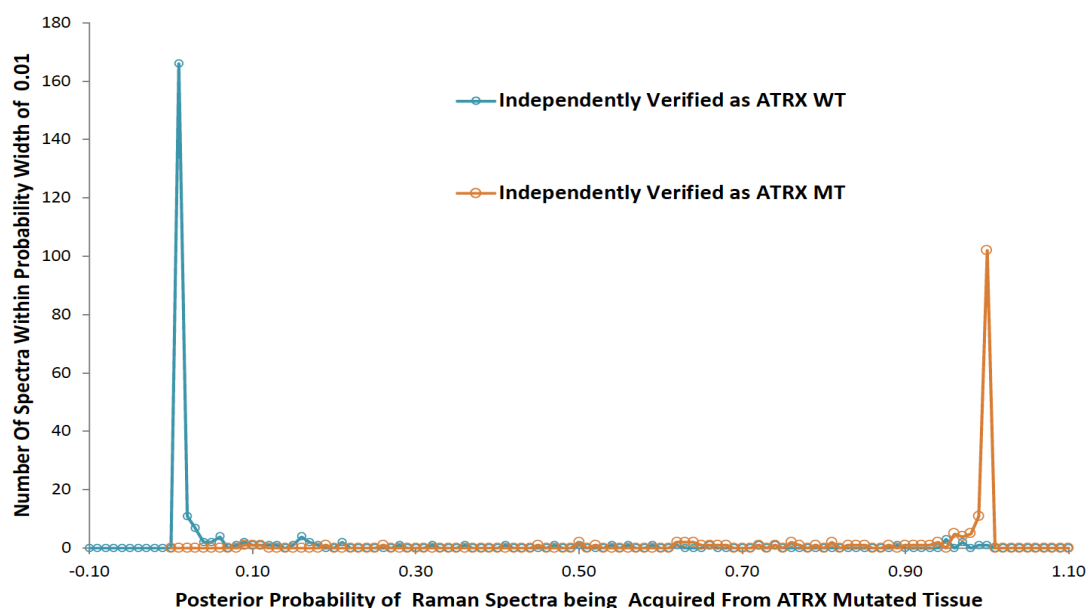


Figure 103. Frequency distribution in the probability of brain tissue having the ATRX mutation using PCA-LDA.

A clear separation between ATRX WT (blue) and MT (orange) as identified by histopathology is seen, with little overlap.

The ROC curve in Figure 104 shows the excellent performance of this model with leave 20% out cross validation giving an AUC of 99% and a sensitivity and specificity of 96%.

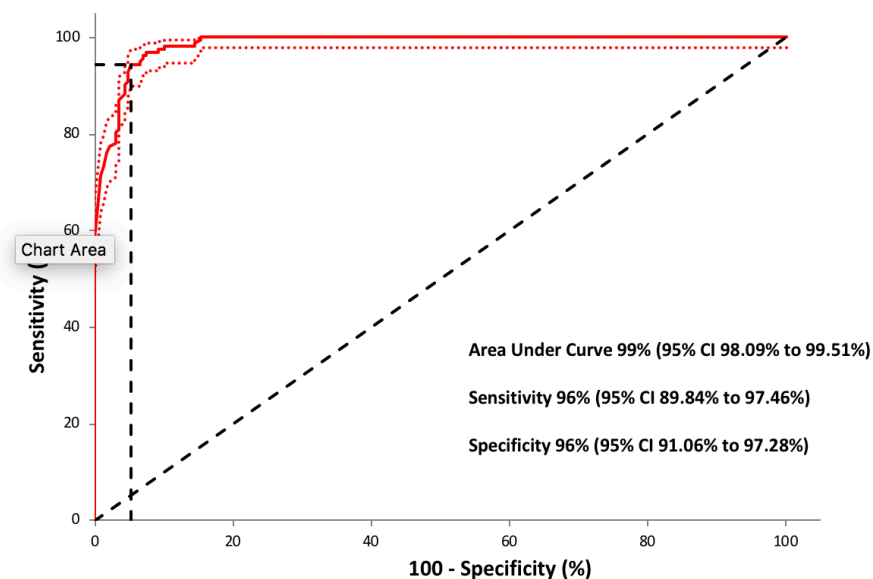


Figure 104. ROC curve and AUC for the PLS-LDA model when predicting ATRX mutation.

With an AUC of 99% and a sensitivity and specificity of 96% the PCA-LDA model shows excellent ability to predict the ATRX status of tissue being analysed.

SVM and PCA-LDA produced the best predictive models for ATRX mutation with high sensitivity and specificity.

4.7.10.4. MGMT

Average spectra MGMT Methylated and MGMT Unmethylated tissue are shown in Figure 105. The difference spectrum in Figure 105 shows many areas of significant difference as calculated using the Mann-Whitney U test for non-parametric data.

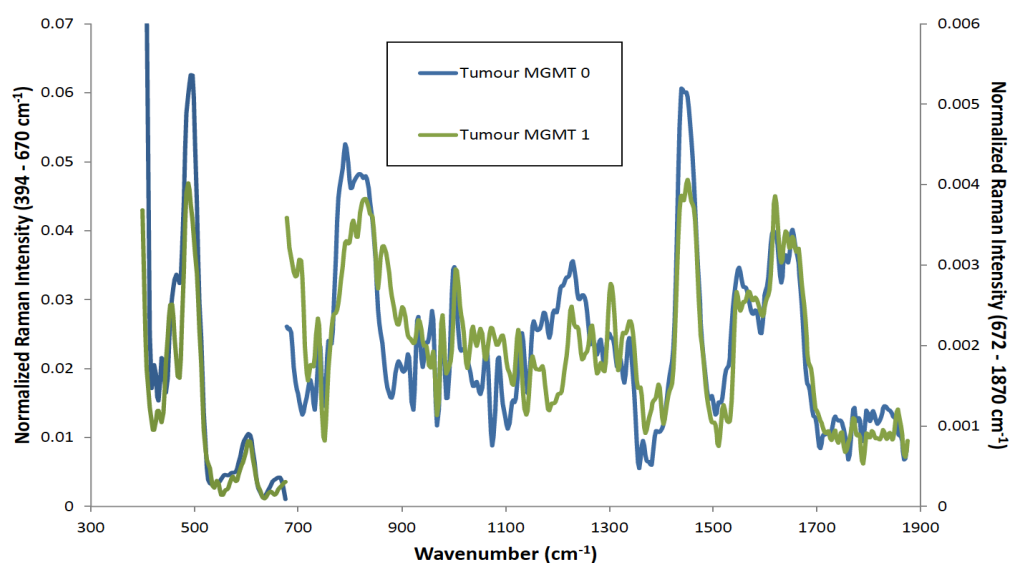


Figure 105. Average spectra from MGMT methylated and MGMT unmethylated tissue.

Spectra from MGMT methylated (green) and MGMT non-methylated (blue) tissue are shown superimposed in the 390-1800 wavenumber range. The break in spectra at 690 is due to the optical filters used in the system. Differences are visible throughout the DNA and protein and lipid regions of the Raman spectrum.

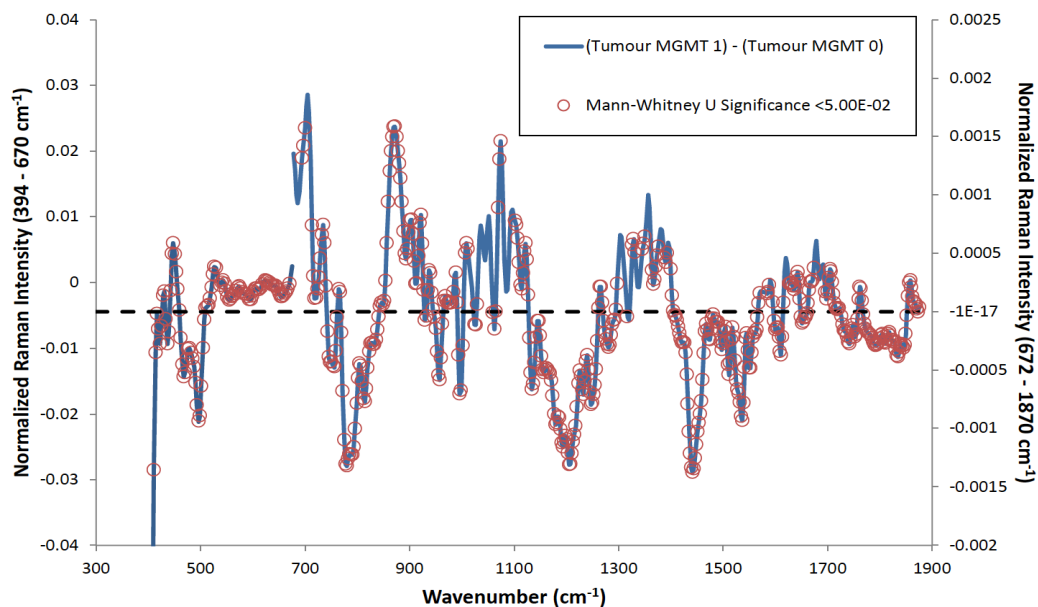


Figure 106. MGMT Raman difference spectrum.

Areas with statistically significant difference according to the Mann-Whitney U test are identified with red circles, $P < 0.05$.

MGMT difference spectra grouped by grade of tumour and other mutations are shown in Figure 106 forming the basis for an attempt to identify reference spectra which may be associated with the differences between MGMT methylated and non-methylated tissue spectra.

of variables that go into making these difference spectra (tumour type, ATRX and IDH status) are reduced using hierarchical supervised clustering as shown in Figure 106 many of the DNA and phenylalanine peaks being to emerge (vertical dashed lines) as shown in Figure 106 suggesting a possible molecular cause for the differences being observed.

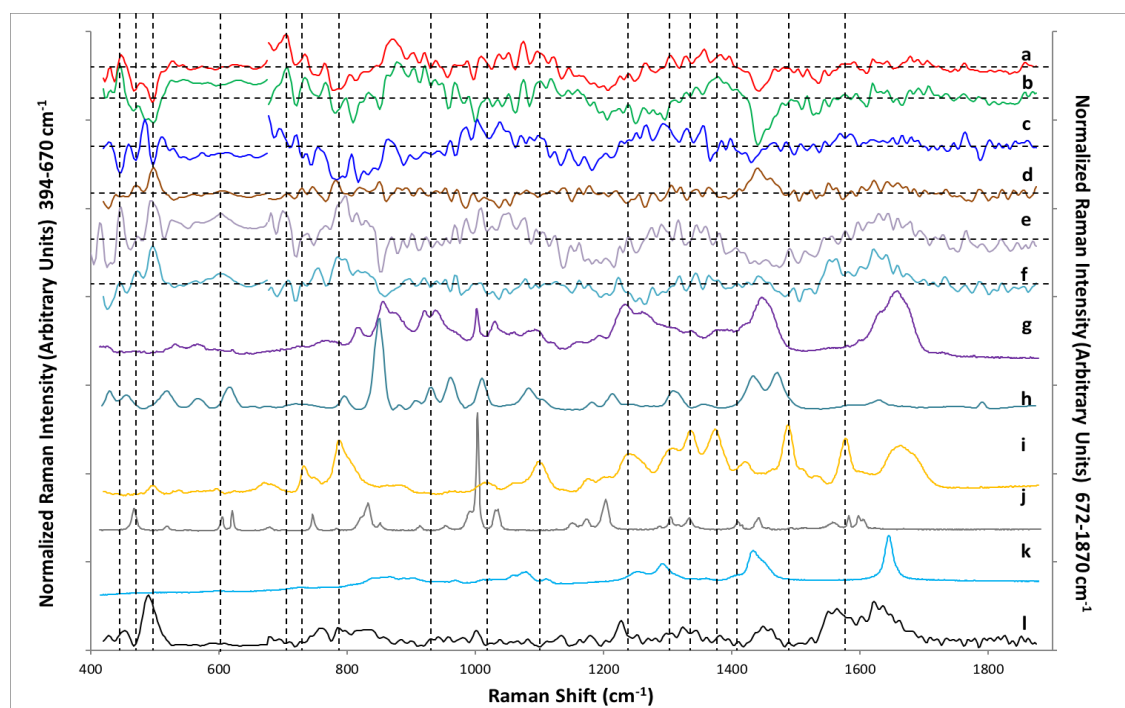


Figure 108. Comparison of successive difference spectra for MGMT methylation status with reference spectra.

a) all 314 tumour spectra tested for MGMT mutation; b) all 138 G2 tumour spectra tested for MGMT mutation; c) all 173 G4 tumour spectra tested for MGMT mutation; d) all 69 G2 tumour spectra tested for MGMT mutation and with ATRX=1 and IDH=1; e) all 54 G4 tumour spectra tested for MGMT mutation and with ATRX=1 and IDH=1. f) all 147 G2, G3 and G4 tumour spectra tested for MGMT mutation and with ATRX=1 and IDH=1. g) collagen; h) 2-HG; i) DNA; j) phenylalanine; k) Lipid (oleic acid); l) human blood.

Unsupervised analysis using PCA shows a good separation between methylated and non-methylated tissue spectra as shown in Figure 108. This suggests classification models should perform well when asked to differentiate between the 2 groups of spectra.

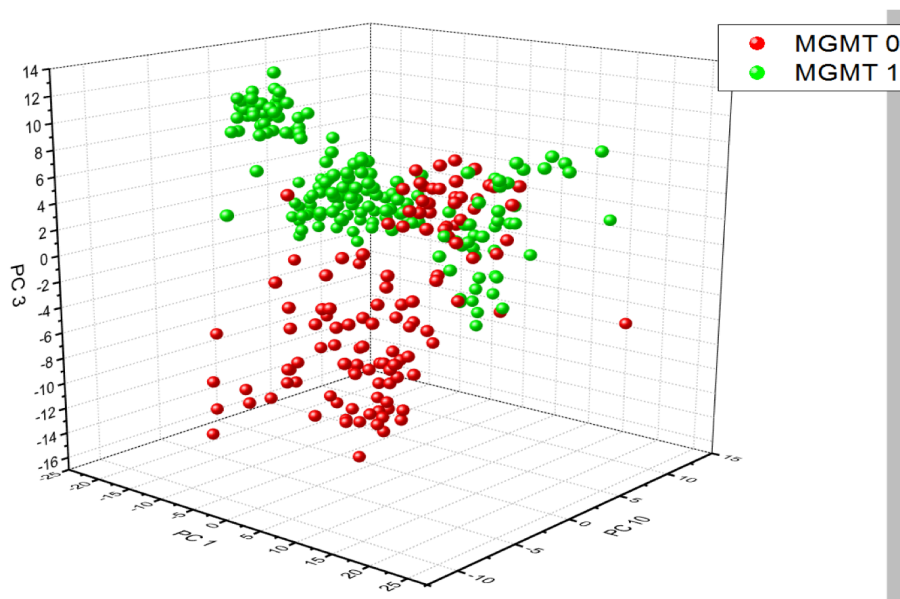


Figure 109. PCA plot of Raman spectra from MGMT methylated vs non-methylated tissue.

There is good segregation between MGMT methylated versus non-methylated spectra with hardly any overlap. MGMT non-methylated =red squares, MGMT methylated= blue circles.

The key peaks responsible for the PCA separation based on my loadings calculations are listed in Table 20a. These represent the most important peaks for discrimination from PC 2,3 and 4 as it was found PC 1 was mainly due to blood. These peaks span DNA, lipid and protein domains.

MGMT	PCA 2	PCA 3	PCA 4
Peak 1	1244 (Amide III)	827 (Nucleic Acid)	1070 (Lipid)
Peak 2	1285 (Lipid)	785 (Nucleic Acid)	996 (Phenylalanine)
Peak 3	1190 (Guanine, Cytosine and adenine)	532 (Cholesterol/protein)	620 (Phenylalanine)

Table 20a. Key peaks seen in principle components 2,3 and 4.

The key lipid, protein and DNA/RNA peaks seen are listed. Peak identity is based on already published literature.

The best classification was again found using RF, SVM and PLS-DA. The RF and SVM accuracy from cross validation is shown below:

Error matrix for the Random Forest model on Rattle MGMT.csv [test] (counts):

Predicted

Actual	0	1
0	66	2
1	7	45

Error matrix for the Random Forest model on Rattle MGMT.csv [test] (proportions):

		Predicted		
Actual	0	1	Error	
0	0.55	0.02	0.03	
1	0.06	0.38	0.13	

Overall error: 7%, Averaged class error: 8%

Rattle timestamp: 2018-04-15 12:33:44 Babar Vaqas

=====

Error matrix for the SVM model on Rattle MGMT.csv [test] (counts):

		Predicted	
Actual	0	1	
0	68	0	
	3	49	

Error matrix for the SVM model on Rattle MGMT.csv [test] (proportions):

		Predicted		
Actual	0	1	Error	
0	0.57	0.00	0.00	
1	0.02	0.41	0.06	

Overall error: 3%, Averaged class error: 3%

Rattle timestamp: 2018-04-15 12:33:45 Babar Vaqas

=====

My SVM and RF classifiers performed extremely well as shown in the sensitivity/specificity plot in Figure 109. With a sensitivity of 94% and a specificity of 100% the SVM model performs the best.

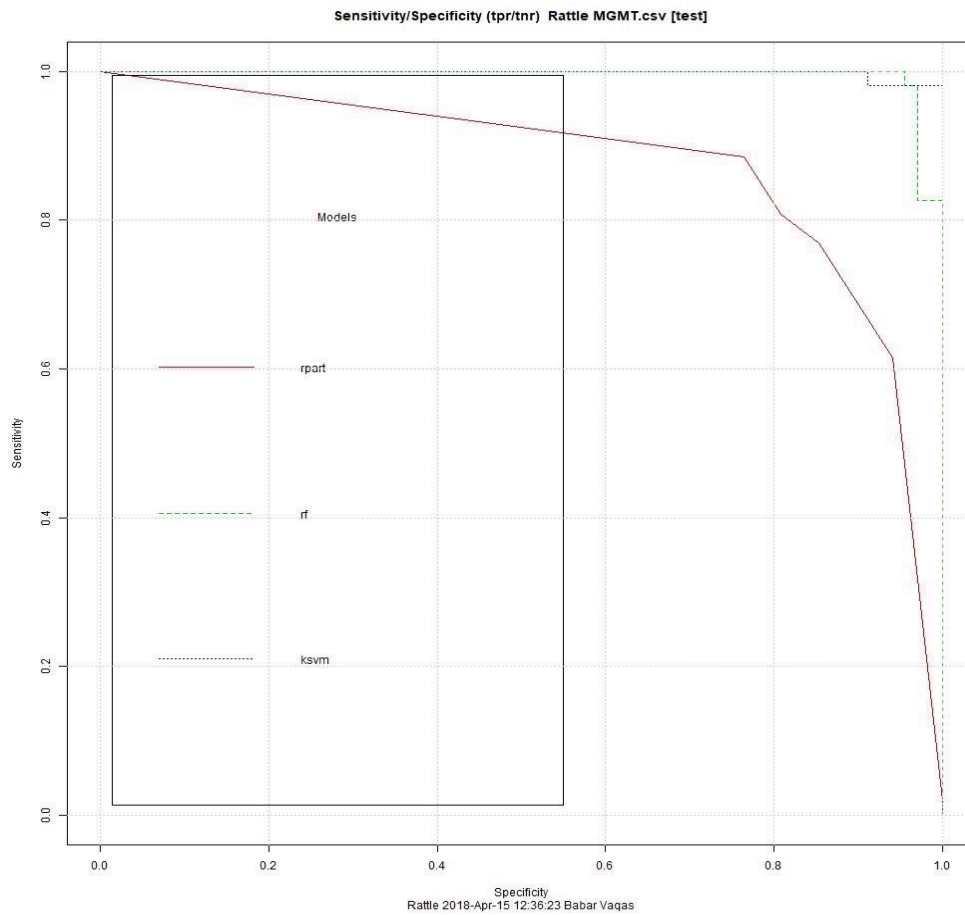


Figure 110. Sensitivity/Specificity plot comparing SVM and ensemble methods model performance for MGMT status

Superior performance of SVM and RF is shown (labelled as ksvm and rf respectively on the chart).

The ROC curve for SVM is excellent, showing an AUC of 1 as shown in Figure 111.

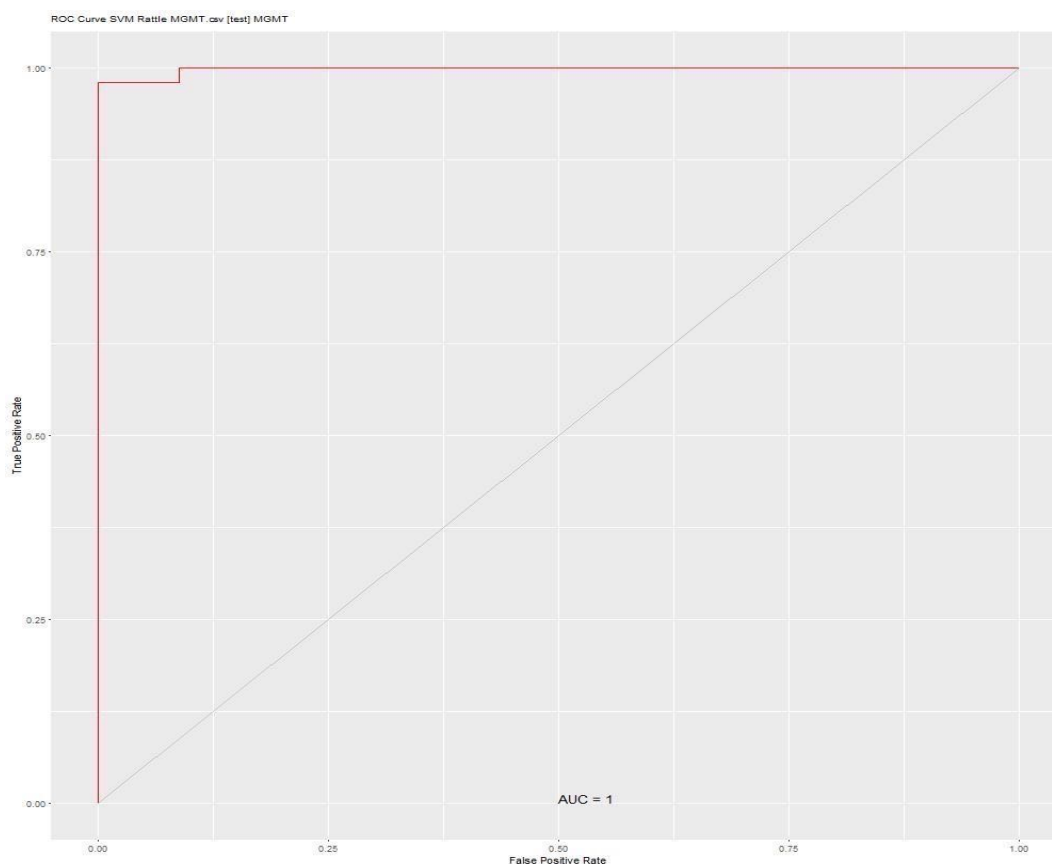


Figure 111. ROC curve and AUC for the PLS-LDA model when predicting IDH mutation.

With an AUC of 100% SVM model shows excellent ability to predict the MGMT status of tissue being analysed.

PCA-LDA models also performed very well, with some of the best separation seen in any of the mutations tested. When looking at the posterior probability distribution, 2 clear populations of spectra are seen in Figure 112 and Figure 307113 with almost no overlap showing that the classifier works well.

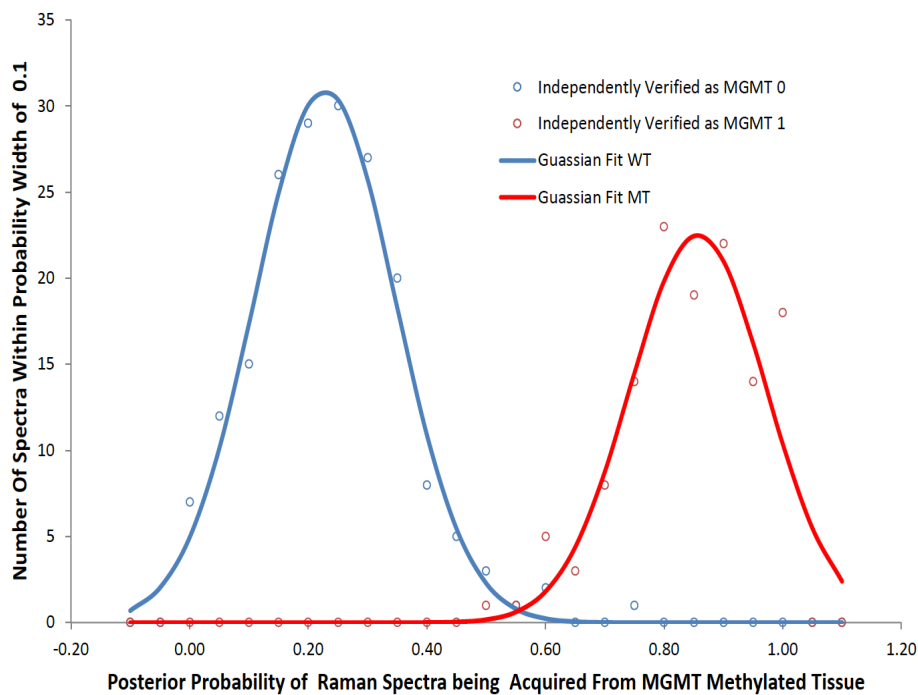


Figure 112. Frequency distribution in the probability of brain tissue having the MGMT methylation using PCA-LDA.

A clear separation between MGMT non-methylated (blue) and MGMT methylated (red) as identified by histopathology is seen, with little overlap.

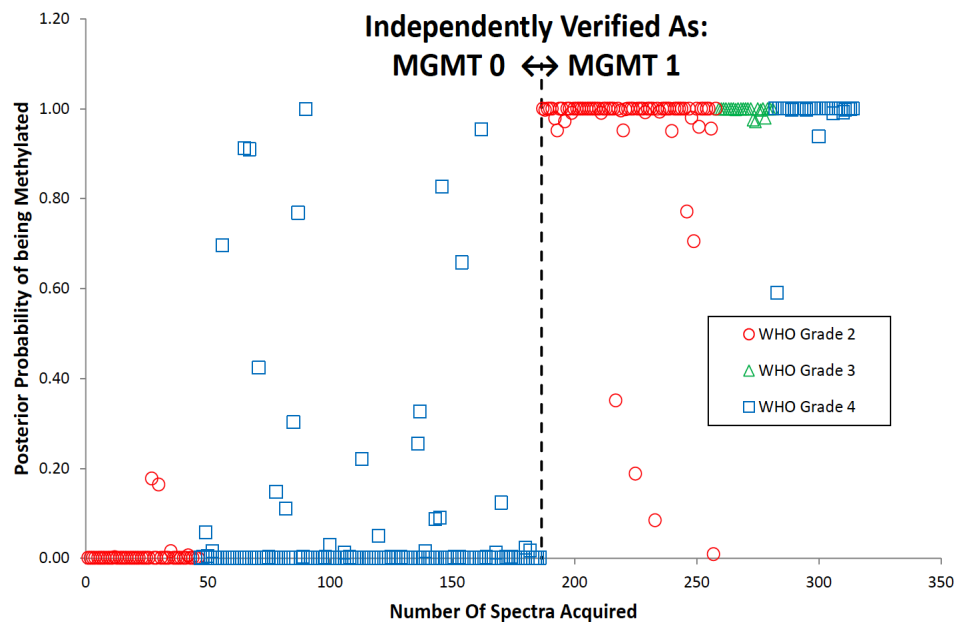


Figure 113. Probability of MGMT mutation in brain tissue using PCA-LDA prediction models.

Good separation between MGMT methylated and non-methylated status tissues is seen with good separation for all WHO Grades.

The ROC curve in Figure 114 confirms excellent classifier performance with an AUC

of 99% and 96% sensitivity and specificity.

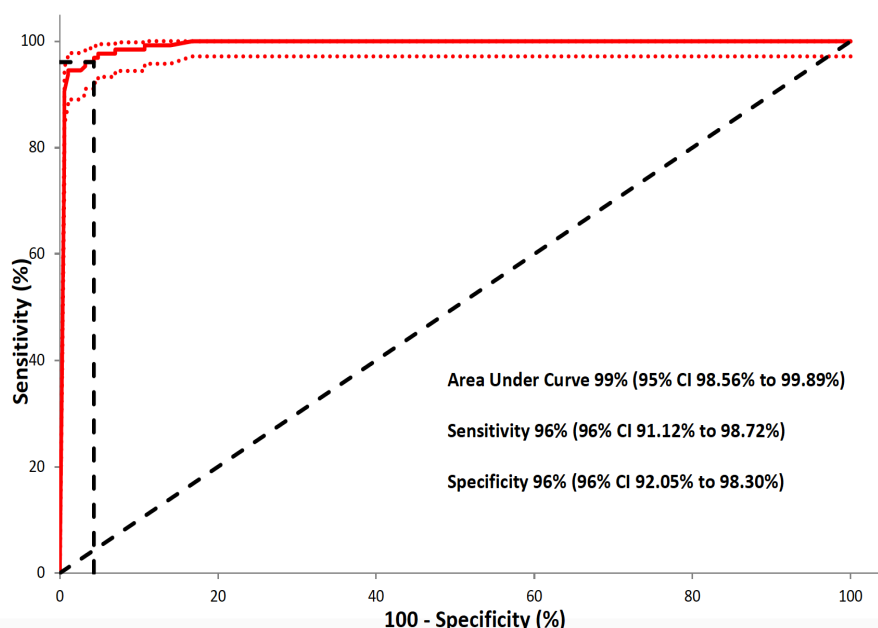


Figure 114. ROC curve and AUC for the PLS-LDA model when predicting MGMT status.

With an AUC of 99% and a sensitivity and specificity of 96% the PCA-LDA model shows excellent ability to predict the MGMT status of tissue being analysed.

4.7.10.5. Summary

The spectra analysed show significant differences between mutated and non-mutated tissue in terms of average spectra and PCA and LDA analysis. The average difference spectra can be shown to have a relationship with peaks from DNA in the case of ATRX and MGMT mutations and 2-HG reference spectra for IDH mutations. Although this is not proof, this does represent maximal use of my current dataset and is a basis for further work.

For all genetic mutations PLS-DA analysis performed the best, with SVM and RF performing slightly inferiorly. Table 21 summarises the range of sensitivities and specificities obtained using PLS-DA.

Status	Sensitivity (%)	95% CI for Sensitivity	Specificity (%)	95% CI for Specificity	AUC ROC (%)	95% CI for AUC
IDH Mutant	94.2	90.6-96.7	94.2	89.5-97.2	98.4	97.3-99.5
ATRX Loss	95.1	90.6-97.9	96.1	92.7-98.2	99.1	98.0-100
MGMT Methylated	98.4	94.5-99.8	98.9	96.2-99.9	99.8	99.6-100

Table 21. Partial Least Squares/Linear Discriminant Analysis (PLS-LDA) performance using Raman spectroscopy for the prediction of key genetic and epigenetic changes in glioma.

Sensitivity, specificity and Area Under Curve (AUC) for the Receiver Operating Characteristic (ROC) was calculated for each type of change observed. PLS-DA classified IDH and ATRX mutations and MGMT promotor methylation with excellent sensitivity and specificity, and a high AUC, demonstrating excellent classifier performance. Abbreviations used: IDH= Isocitrate Dehydrogenase gene, ATRX=Alpha Thalassemia/Mental Retardation Syndrome X-linked gene, MGMT=O6-methylguanine DNA methyltransferase.

This represents a high degree of accuracy for detecting tissue containing these genetic changes which drive glioma formation.

4.7.11. Conclusion

The data I have presented in this chapter shows that Raman spectroscopy when used *in vivo* can provide recognisable Raman spectra. These spectra can be used to show differences between tissue types to answer key clinical questions. I have shown these differences can be detected using averaged spectra and supervised and unsupervised analysis. I have successfully trained machine learning models on the data obtained. These models can be successfully used to predict the clinical characteristics of data newly used for cross-validation to find out if tissue samples are normal or tumour, the type of tumour, tumour grade, tumour density and the genetic mutational status of tissue being analysed. RF, SVM and PLS-DA methods achieved the best results with high accuracy in this study.

Analysis of this data has shown that using these models, it is possible to analyse tissue at the end of surgery at the resection margin and successfully predict the presence of tumour and normal tissue in low grade glioma.

5. Rapid Evaporative Ionization Mass Spectrometry (REIMS)

5.1. Chapter Overview

This section deals with the development and application of REIMS as a new *in vivo* technique for brain surgery. I overcame a number of challenges before using the system in humans including system development, calibration and setup in the operating room. The development of my specific brain tissue data analysis methodology will be discussed. The results of the first *in vivo* use of the system will be presented using multivariate statistics. I aim to show how a laboratory technique can be modified for use in an environment such as an operating theatre and yet retain enough accuracy for important conclusions to be drawn from the tissue being analysed.

5.2. Method Development

5.2.1. Introduction

Development of a method which enabled mass spectrometry readings to be taken during surgery which were repeatable and reliable was important. Method development consisted of the systematic identification and control of modifiable factors contributing to the unwanted variation in mass spectra recorded. The aim of method development was the creation of an iKnife surgical protocol for this and future mass spectrometry trials which can increase the yield of usable mass spectra recorded during surgery. Approaching this problem systematically the system can be physically broken down into its components - the bipolar forceps, the tubing connecting the forceps to the mass spectrometry machine, the choice of solvents used on surgical vapour prior to entry into the machine and the location of the machine itself.

Method development was rapid as recruitment time for the study was limited due to cost, operating and hospital environment constraints. Hence I used the most efficient and fastest approaches to control variables. Methods and results will be presented together as my iKnife method development for brain surgery. Method development was an ongoing process starting before the system was deployed in theatre and continued into the *in vivo* phase as the system was used in the first 36 patients.

5.2.2. Instrument location

Instrument location must be safe for the patient and staff in the operating theatre. It must also be optimal to allow for the collection of as much mass spectrometry data as possible in one sitting. Data is unlikely to be collected in enough quality or quantity if using the machine introduces unnecessary delay in operating- either through the need to move the machine each time, or if the location far away from the patient sacrifices data quality. Factors affecting location include:

1. Safety: machine ventilation (overheating), power leads and air supply tubing (falls risk), obstruction to other machines used for surgery, passing of instruments to surgeon
2. Noise: 3-4 hour operating time increases sensitivity to noise of surgeon and theatre staff. In some cases, vacuum could not be produced as the machine had to be switched off due to noise if too close to the surgeon.
3. Bipolar suction tubing length: longer tubing increases the time delay to obtaining a mass spec reading. It also makes it more difficult to localise tube blockage from blood/saline as there is more tubing to inspect: this reduces the data collection ability of the system.

After discussion with surgeons and a trial period of 2 weeks, I determined the ideal location for the machine in the operating theatre. All 3 surgeons using the operating theatre had input into this, with patient safety and surgeon comfort being the most important criteria. Instrument safety dictated the machine should be in a well ventilated area to prevent overheating and cables should not pose a trip hazard, it therefore needed to be close to a power and air supply. The ideal location was determined to be on the surgeon's left hand side (not impacting on surgical equipment passed by the scrub nurse to the surgeon's right hand). An area near the periphery of the operating room was ideal as it was close to a suspended power rail and a high pressure air connection point avoiding long cables which could pose a trip hazard to staff. A suitable distance (3-4 metres) was agreed as far enough from the patient to allow other machines to be positioned (operating microscope and drills) and to avoid problems due to vacuum noise of the machine (for the comfort of the surgeon) but a short enough distance not to cause an impractical delay in mass spec readings (3-5 second delay achieved from triggering bipolar foot switch to first readings on mass spec machine). The final position of the machine is shown in Photograph 4.



Photograph 4. The layout of Theatre 7, Charing Cross Hospital.

The operating table is in the centre of the operating theatre (green gel mat is seen on top) with surgical equipment deployed radially around it. The patient is positioned with the head at the end of the table which is furthest from view. From left to right: Sonowand, Raman system, iKnife next to the wall. This was the optimum position for the machine in terms of closeness to the patient and being on the left hand side of the patient, avoiding obstruction to other machines and equipment during surgery.

5.2.3. Tubing characteristics

Factors relating to the tubing used to connect the bipolar forceps to the mass spectrometry machine included the material the tubing is made out of, its length and diameter. Functional concerns included long tubing being a falls risk (safety concern), longer tubing increasing the time taken for surgical vapour to reach the mass spectrometry machine and a narrow diameter resulting in frequent blockage of the tubing by fluid aspirate from the operative field preventing any further data collection until the blockage was cleared. This may result in the machine being unavailable for data collection while it is being unblocked, resulting in loss of data. Hence the optimal tubing length, diameter and material needed to be defined for the best data collection.

I compared tubing material from Erbe and Hygitech by running IPA solvent only to see if the tubing itself had a signal as shown in Figure 115.

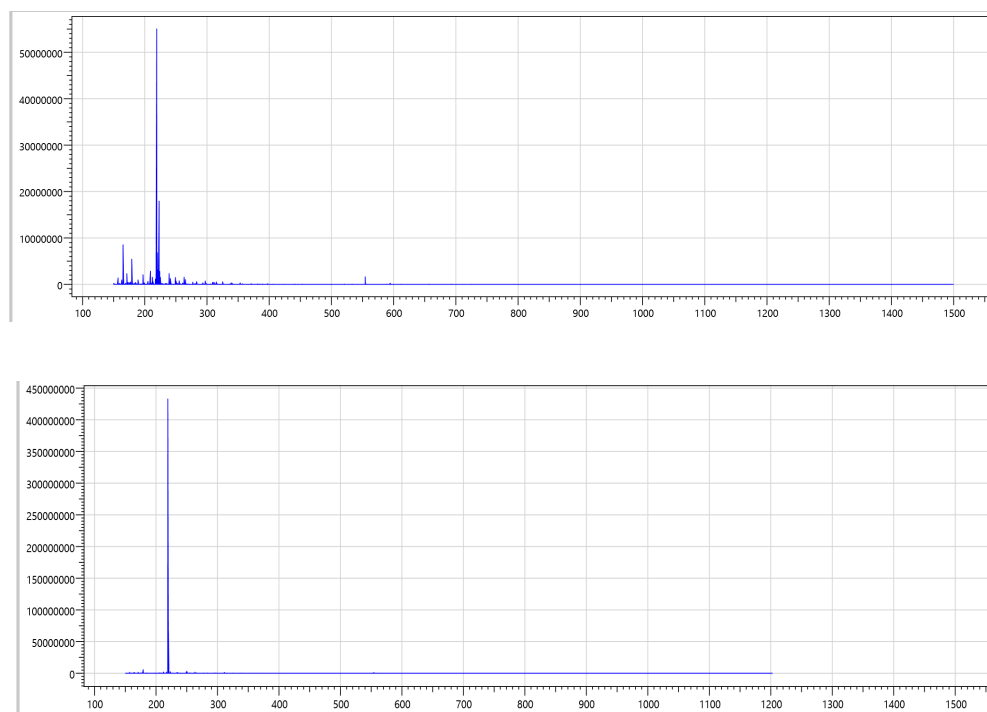


Figure 115. Mass spectra of ERBE versus Hygitech tubing.
Hygitech tubing (above) and Erbe tubing (below) was compared to look for effects on REIMS spectra.

The results show the IPA signal as being the highest intensity signal in the lower mass range. Interestingly, the Hygitech tubing has multiple low masses in the spectrum which create unnecessary noise.

Erbe tubing contains no such noise. It was decided to use the Erbe tubing for iKnife experiments to keep noise to signal ratio acceptable.

In order to investigate the remaining factors, the following setup was used. Calf brain tissue was repeatedly sampled in the iKnife laboratory using a Xevo mass spectrometry machine (the same machine used in the operating theatre). IPA was

used as the solvent, multiple tissue burns were created with the generator set at 20 and 2 seconds contact time with tissue. Tubing diameter was found to have a negligible effect on signal but a significant effect on chances of blockage in practice, hence the largest standard diameter fitting Erbe irrigation bipolar was used. Tubing length was found to have a negligible effect on signal and signal time, hence shortest length used to prevent falls risk (2.8m).

5.2.4. Bipolar type

2 irrigation bipolar forceps were compared: Braun vs Erbe. Both forceps can be connected to the mass spectrometry machine but have different designs which could influence the observed spectra. Both forceps were used in the laboratory on pork liver to assess which one produced the best spectrum, as shown in Figure 116.

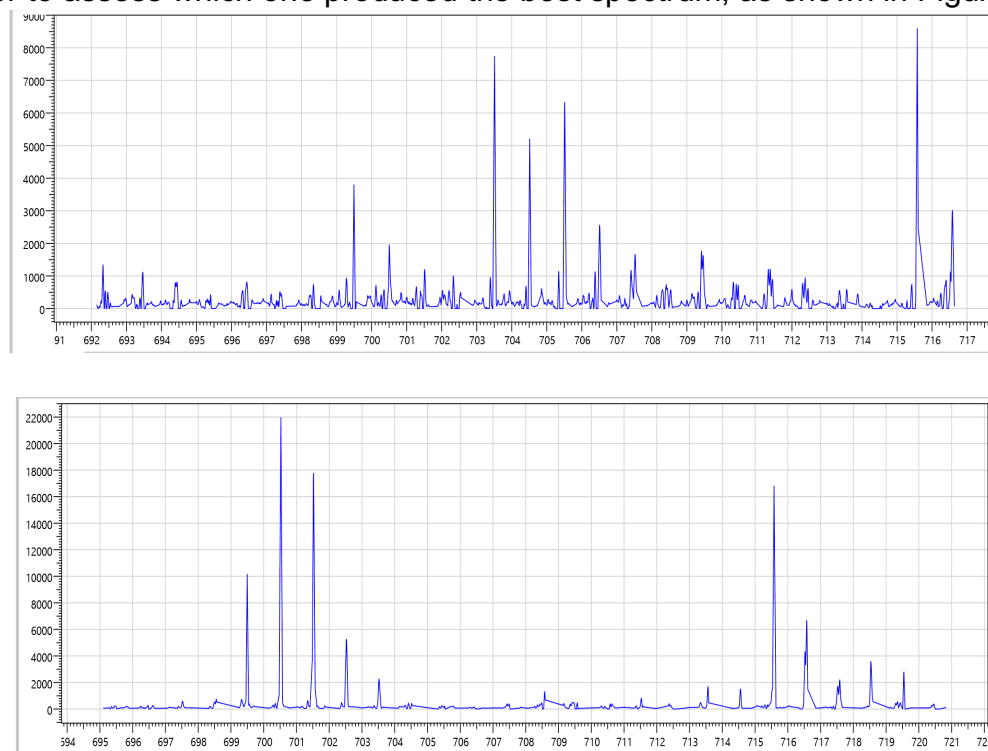


Figure 116. Above: Mass spectra obtained using Braun forceps. Below: Mass spectra obtained using Erbe forceps.

Erbe forceps seem to provide a better signal intensity and better signal to noise ratio. This may be because the Braun forceps had the irrigation hole very high up in the handpiece, further away than the Erbe handset which had the opening for irrigation at the tips of the bipolar forceps (figure 117). The disadvantage of having the opening at the forceps tip is that blockage of this aperture can occur more easily than in the Braun design.



Figure 117. Erbe Bipolar irrigation forceps.

ERBE Elektromedizin GmbH, product code 20195-150. The irrigation hole is at the forceps tip, resulting in a cleaner, higher amplitude signal for suction of surgical vapour into the mass spec system.

It was decided to use the Erbe forceps shown in Figure 117 for in vivo work.

5.2.5. IPA vs no IPA

The first 10 patients exhibited significant noise in the mass spectra. In order to investigate methods to improve this Isopropyl Alcohol (IPA) was used as a solvent for the aspirated smoke from bipolar use.

The IPA peak is at around 220 on the x-axis. Initial testing of IPA vs Non-IPA was on calf brain in the laboratory and suggested significant differences between spectra, as shown in Figure 118.

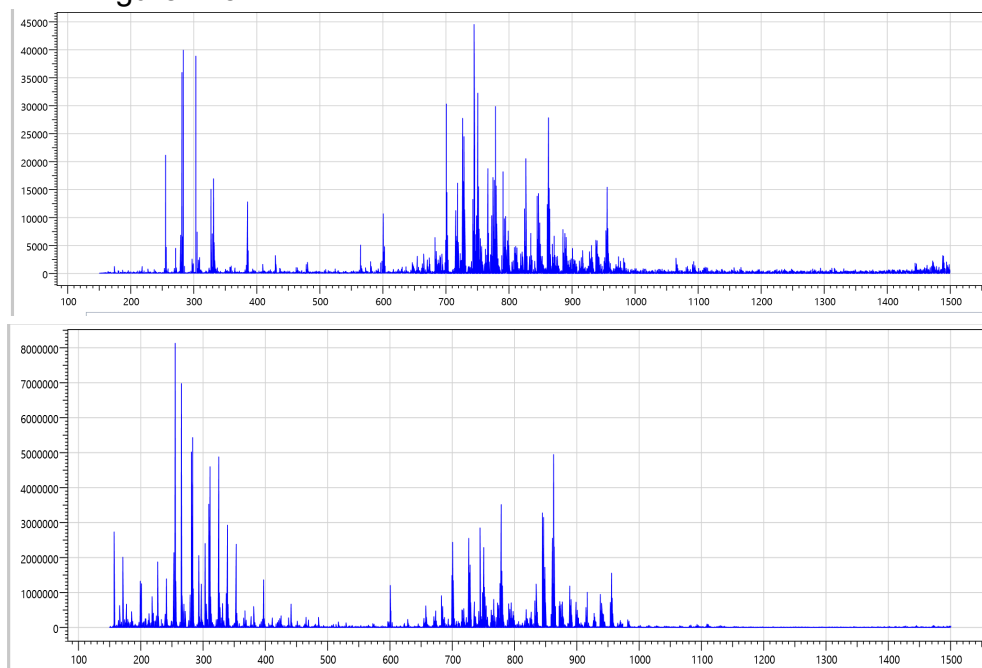


Figure 118. Calf brain REIMS spectra illustrating the effect of IPA.

Non-IPA MS of brain tissue (top) in negative mode shows a noisy baseline as seen in at 1500 m/z value. Using IPA (bottom), noise is significantly improved almost to zero, whilst preserving most peaks. Peak intensity is reduced slightly, with some change in peak morphology.

Testing was then done on ex vivo brain tissue samples which revealed similar results to calf brain tests as shown in Figure 119. The results were much better than with the

calf brain model suggesting significant improvement in accuracy using IPA. The measured intensity also increased.

Signal in m/z 600-1080 range – Brain ex vivo

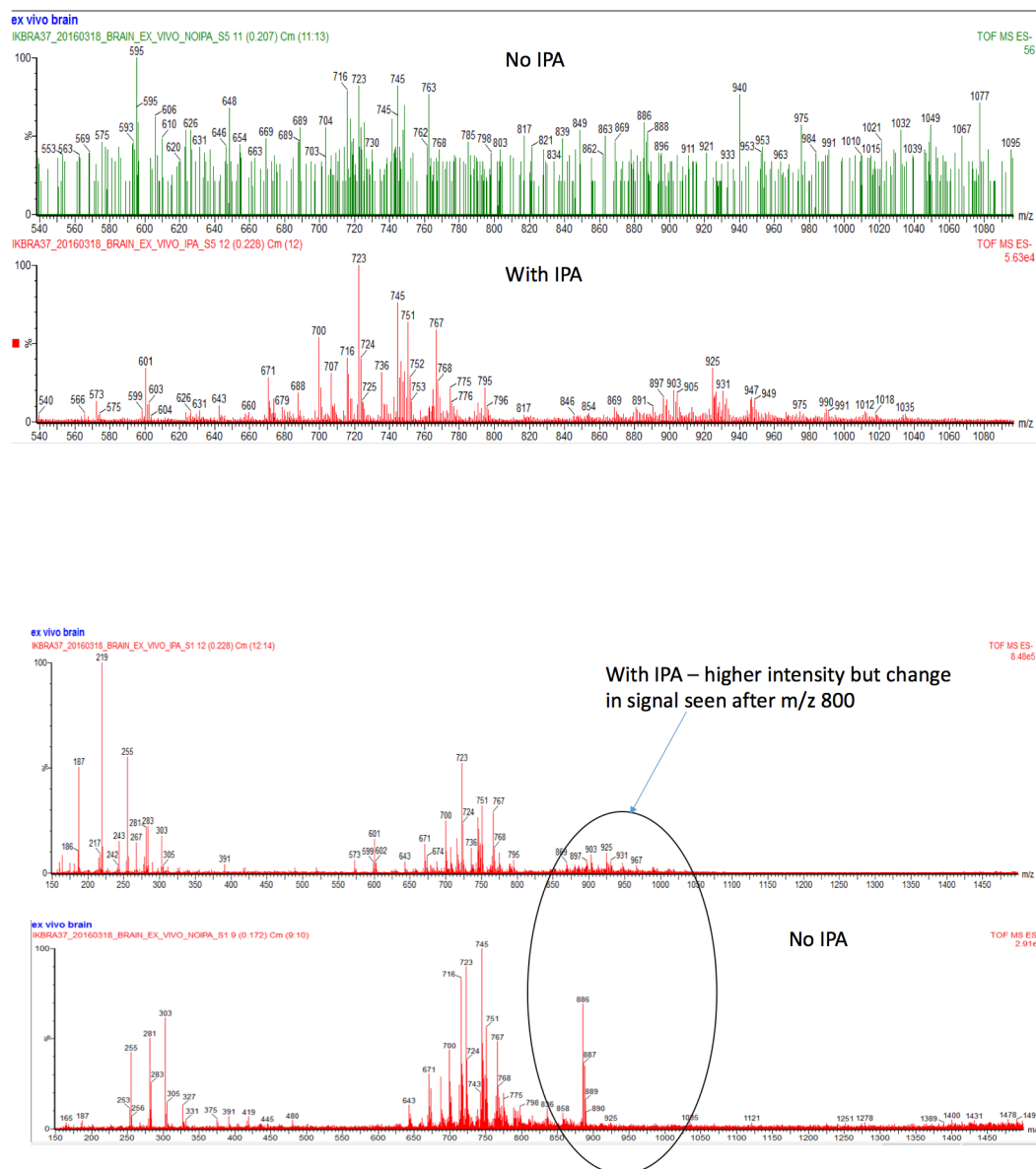


Figure 119. ex vivo testing of IPA using brain tumour tissue from patient 37.
Top: Spectra were acquired without (green) and with IPA (red) with clear differences in spectral intensity, noise and peak morphology.
Bottom: IPA spectra are of higher intensity, with less baseline noise and have a changed morphology particularly after m/z 800.

In order to see if IPA was the best solvent other alcohols were tried in the lab: methanol and pentanol as shown in Figure 120 below.

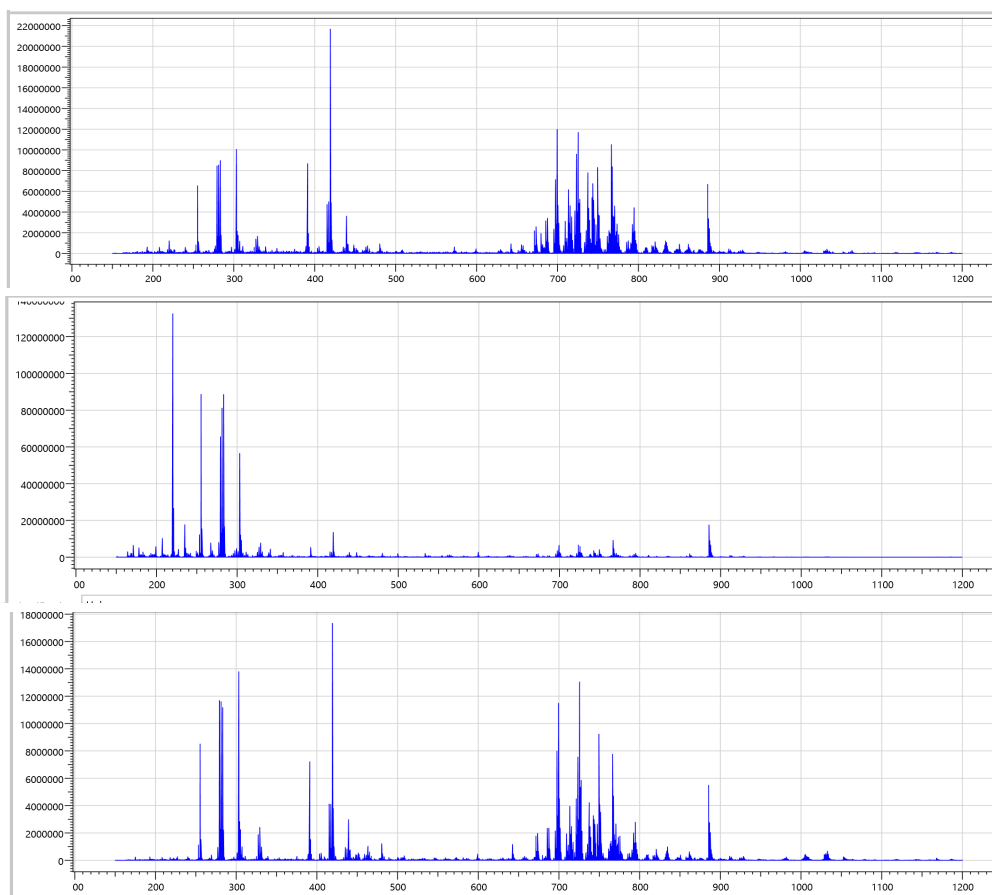


Figure 120. ex vivo testing of alcohol solvents on pork tissue.

Mass spectra from pork liver obtained in the laboratory using IPA (top), methanol (middle) and pentanol (bottom). A range of mass spectral morphologies are seen with the different solvents.

The best alcohol solvent for mass spec keeps signal to noise ratio low and results in a more spread out peak pattern as it aids fragmentation of molecular clusters, increasing the amount of useful information gained in terms of profiling tissue. A range of alcoholic solvents were trialled including Isopropyl Alcohol, Methanol and Pentanol on pork liver tissue (Figure 120). IPA was chosen because it resulted in most information in the 600-900 range which has shown to be most important for tissue profiling in ex vivo brain samples.

5.2.6. Effect of Blood on the spectrum

Brain tissue and brain tumours are quite vascular. Bleeding is often encountered when operating and the presence of blood was notable during data collection with iKnife in the first cohort of patients. This was a problem for a number of reasons. In some cases, blood obscured the operative field and the iKnife sampled blood rather than tumour tissue. The resultant spectrum contained molecules in the lower mass ranges, as shown in Figure 121, as blood is not as phospholipid rich as brain tissue. In other cases, a mixed blood/tumour spectrum was obtained with information in the lower and higher mass ranges.

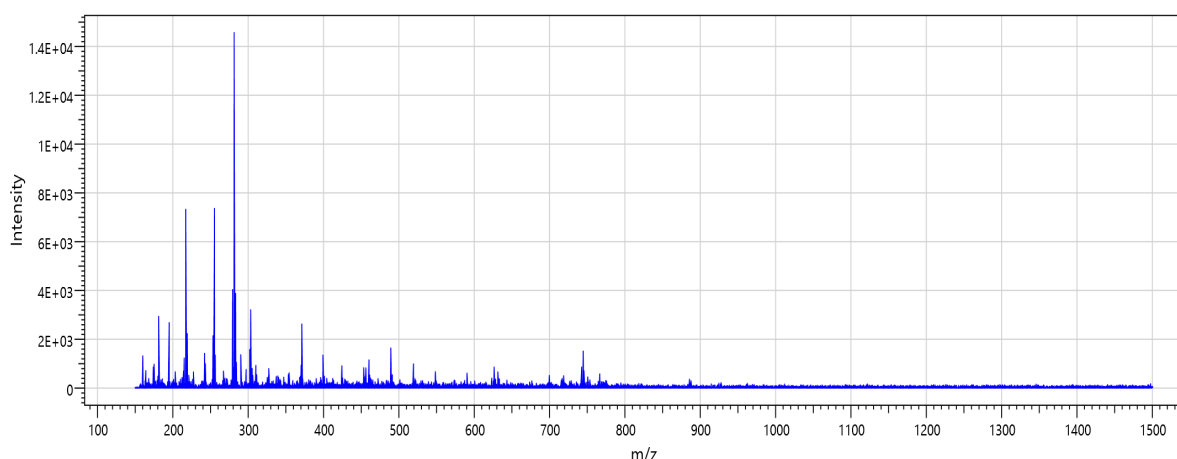


Figure 121. Mass spectrum obtained when sampling fresh haematoma forming during surgery.

This spectrum contains information mainly in the lower mass range which is not diagnostically useful.

Hence in cases where the majority of the information is in the lower mass range, adequate haemostasis at the operating site was always checked.

5.2.7. Conclusions

I investigated a number of variables leading to the final setup for *in vivo* data collection. Erbe tubing and bipolar forceps were used with a minimal safe tubing distance and as large a diameter as possible to prevent blockage. Tissue was carefully haemostased prior to iKnife readings and IPA was used as a solvent for improved mass spectra. In terms of quality criteria for mass spectra the following were used:

1. Baseline as flat as possible (low noise)
2. Ensure m/z 600-1000 range has data (diagnostic part of spectrum)
3. Ensure adequate haemostasis when obtaining spectrum

If these criteria were not met the reading was repeated.

5.3. Clinical Trial Protocol

5.3.1. Introduction

This section deals with the REIMS/iKnife clinical trial protocol that was developed. Detailed information on instrument setup, calibration and use during surgery is presented. The aim is to outline a structure for intraoperative sample annotation and chemometric analysis.

5.3.2. Instrument setup, Calibration and Troubleshooting

Instructions for system setup, calibration and troubleshooting are included in Appendix 3, section 8.3.

5.3.3. Collecting *in vivo* data: Surgical Sampling

In most cases I collected iKnife data with Raman spectroscopy as has already been mentioned in section 4.5. Some cases were iKnife only (if sterile Raman equipment was not available). For all iKnife cases the following steps were taken to obtain and record data:

- Identification of suitable point for mass spectrometry
- Verbal check with scientist that iKnife is ready to analyse data
- Erbe irrigation forceps checked for blockage: flush tubing if blockage found
- Mass spectrum reading (settings and setup as in method development)
- Location logged on neuronavigation
- Biopsy taken
- Biopsy labelled anatomically and numerically by scrub nurse, dictated by BV
- iKnife data acquisition file named numerically
- Specimen label checked with iKnife records
- Specimen sent to histopathology for immediate reporting (fresh) and definitive histopathology (see section 4.5.4).

Figure 122 shows the bipolar irrigation forceps in use in order for iKnife data to be obtained. Irrigation tubing starting at the end of the blue bipolar handpiece connects to larger diameter suction tubing to connect with the atmospheric interface of the REIMS machine. Blockage can occur from the bipolar tip aperture to anywhere in the tubing. Suction pressure can be measured at the atmospheric REIMS interface which can suggest tube blockage.



Figure 122. The use if the iKnife during brain surgery.
Irrigation bipolar forceps are shown connected to irrigation tubing which connects to the REIMS atmospheric interface.

REIMS spectra are obtained throughout the operation. As shown in Figure 123, spectra are taken at the brain surface, tumour core to the resection margin.

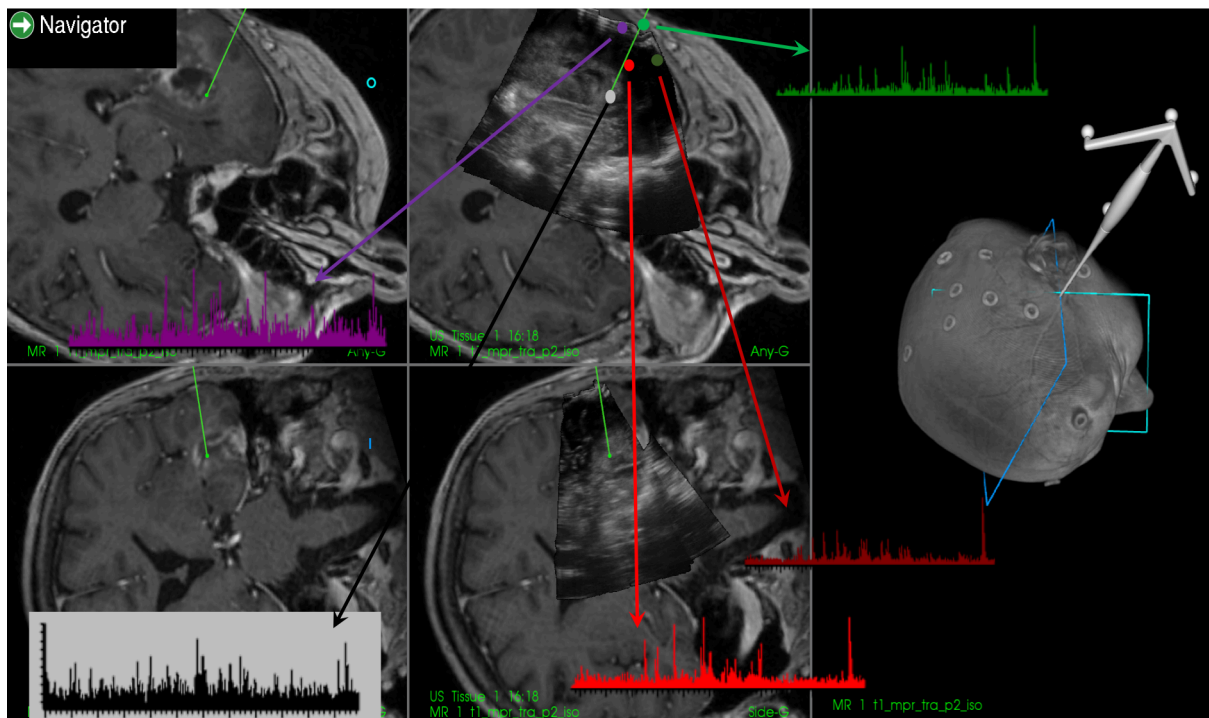


Figure 123. Progressive sampling during an iKnife case.

Each spectrum (superimposed onto the combined MRI/Ultrasound image) was registered with a point in 3D space using Sonowand neuronavigation. The points where the spectral information was taken are shown as coloured dots, with data being obtained deeper into the tumour as surgery progresses.

5.3.4. Pathology Specimen Processing

Protocol for collection and storage of fresh and frozen tissue for REIMS analysis:

- Surgical staff will immediately inform the pathology team of the arrival of a new specimen and a member of the portering staff/research team will collect it from the theatre.
- The surgical specimen together with the histopathology request form will be delivered from the operating theatre by the portering staff/Clinical Research Fellows (CRFs) and taken to Pathology Department where it will be received by the appropriate pathologist or his/her nominee.
- The surgical specimen will be delivered by 30 minutes from surgical removal. If necessary, keep tissue hydrated by wetting with distilled water.
- Otherwise, if portering staff/CRFs are not immediately available, fresh samples will be stored in the theatre in a fridge at +4 °C for up to 4 hours.
- Fresh samples for research will be taken at the pathologists' discretion, where this does not compromise the diagnostic assessment.
- In the event that the lead pathologist or his/her nominee are not available to take or assess samples from surgery, no material will be taken for research under any circumstances.
- No fresh samples for research will be taken from operative specimens from patients where the presence of a communicable infection is known.
- The pathologist or CRFs will place the tissue, provided by the pathologist for research, in an Eppendorf tube or in a small tube.
- These tubes will be provided by the research team. These tubes will have the expiry date of one week hence labelled on them.
- The sample, received fresh from the pathologist, must be immediately refrigerated and can be stored for up to 4 hours before analysis. After 4 hours it is not recommended to analyse as fresh tissue and must be transferred to a -80°C freezer.
- If the tissue is received frozen, after labelling it must be stored in a -80°C freezer.
- On receipt of a sample for REIMS analysis, it must be assigned a unique sample number appropriate to the project it is for, e.g. for breast tissue this is an IKB number.
- For tissue database purposes, the sample must be logged onto the appropriate project database with unique reference number and storage location.

Genetics analysis was performed by our pathology laboratory and is already described in section 4.5.7.

5.3.5. Summary

I obtained *in vivo* spectra successfully using the trial protocols described in the preceding sections. Successful system calibration and operation resulted in a surgically usable system which did not interfere with surgical workflow in a busy operating theatre. The data collected was analysed using multivariate statistics. This will be detailed in the next section.

5.4. Data Analysis

5.4.1. Introduction

The aim of this section is to describe the REIMS data analysis workflow. I developed a consistent data analysis methodology which works for brain spectra. This is the most time consuming part of the iKnife project. Data must be collected, transferred and manipulated with high fidelity to prevent data loss and failed models. This section is relatively brief as the general principles of chemometric data flow and analysis using machine learning have been covered in section 4.6.

5.4.2. Data sources and integration

Data from a variety of sources was collected for processing. The iKnife machine produced .RAW files containing spectral information. These files were saved on a network drive and were immediately named according to pathology sample number in theatre. The pathology sample number consisted of the IKBK patient number which was unique for every patient. The file also contained a time and date stamp. Pathology samples were labelled with the same patient research code to enable matching to the iKnife spectrum recorded. This was double checked before the samples were sent to the pathology lab.

I collected metadata concerning patient date of birth, gender and operation date. Pathology results were contained on CERNER systems and were accessed on NHS computers. A fully anonymised index file: containing iKnife RAW file names and associated meta data was created. The data fields on this file included the following data for each spectrum:

- Operation date
- Gender
- Patient age
- Histological information (all),
 - Tumour/Normal
 - Glial/Non-Glial Tumour
 - WHO Grade
 - Tumour type
- Tumour cell density (None/Low/Moderate/High)
- Genetics data (IDH, MGMT, ATRX)

The data was then ready for processing in the Offline Model Builder (OMB) from Waters Corporation.

5.4.3. Data pre-treatment

Before a model can be created the mass range which will be used needs to be

decided. Ideally this should be a narrow range containing diagnostic information only, and no noise. This is to ensure the model uses diagnostic information and not noise to base predictions on. There may however be diagnostic information throughout the whole spectrum which may be useful for prediction. Models using a variety of mass ranges were therefore constructed. The broadest range I used was 100-1000, with a limited 500-900 and 600-900 ranges also being used.

The data was normalised using Total Ion Count (TIC) normalisation and de-noised prior to model building using OMB's internal normalisation and background subtraction algorithms. This was essential prior to the application of multivariate statistics so that the variables were measured in terms of ratios rather than absolute numbers (which varied greatly as intensity is measured in arbitrary units only). In addition, all IPA spectra were lockmass corrected to the external compound leucine-enkephalin (negative ionisation m/z 554.2615) contained within an IPA matrix.

A number of options were available to calculate spectra in the recorded RAW file. These include one burn per spectrum, one spectrum per tissue sample and one spectrum per scan. One spectrum per burn was used as the ideal balance between gaining accurate data (individual spectra from tissue) and preventing over fitting in the model by obtaining too much data from a single specimen (one spectrum per scan).

The data was then ready for model building.

5.4.4. Model building

I created models in the Offline Model Builder (OMB version 1.1.29.0) designed by Waters Research Centre, Budapest, Hungary. Individual spectral files were opened in OMB and associated with data labels from the data sources above. An iterative hypothesis driven data exploration approach was used to analyse the data. This used all the data collected, then applied multivariate statistics to discover data structure. Further questions were then developed based on the results of this. A variety of iterations of multivariate statistical models were available and many combinations were tried in order to find the best model for the data obtained. OMB was then used to compute PCA and LDA models of the data. A variety of models were created for a number of comparisons:

IPA vs non-IPA (method development)

in vivo vs *ex vivo*

Tumour vs Normal

Tumour type

Tumour WHO grade

Tumour density

Genomic results: IDH, MGMT, ATRX

For each comparison models were created using a variety of m/z values, PCA only, LDA only and PCA/LDA combination models, IPA, non-IPA and IPA combined with non-IPA data. Cross validation was then performed to assess the performance of the

model, and for some models validation on an external dataset was performed using Recognition software (Waters Research Centre, Budapest, Hungary).

5.4.5. Model performance

Models were tested using cross validation options within OMB. These include 'leave one group out' which is the same as 'leave one patient out' cross validation discussed in section 3.6.7 and leave 20% out cross validation. Leave one out cross validation was unavailable as an option. The resultant confusion matrix was generated as an HTML file which contains information required to calculate the sensitivity and specificity of the model being cross validated. For each cross validation iteration analysis of variance (ANOVA) was used for feature selection ($\alpha < 0.05$). The best performing models were validated on spectra from unseen tissue samples using Recognition software, or from a hold out validation group as part of LOPOCV or leave 20% out cross validation. The results were tabulated as a confusion matrix from which validated sensitivity and specificity was calculated.

5.4.6. Conclusions

This section has detailed my approach to data collection, pre-treatment, model building and assessment of model performance in OMB for iKnife data. The data generated will now be presented in the results sections which follow.

5.5. Results

5.5.1. Introduction

The following chapters present the main results obtained using the *in vivo* REIMS system described in section 3.4. The data will be presented first as a demographic summary of patients and samples from which the data has been obtained followed by a more detailed account of the brain *in vivo* spectra acquired during brain tumour surgery with classification models being presented together with model performance measures. This will provide an overview of the utility of the molecular information obtained using *in vivo* REIMS in brain tumour surgery.

5.5.2. Demographics

5.5.2.1. Introduction

A total of 80 patients were consented for the iKnife study. In 5 cases equipment failure prevented the collection of data, with a resulting 75 patients having data collected for this study. All further statistics refer to this total of 75 patients.

5.5.2.2. Demographic analysis

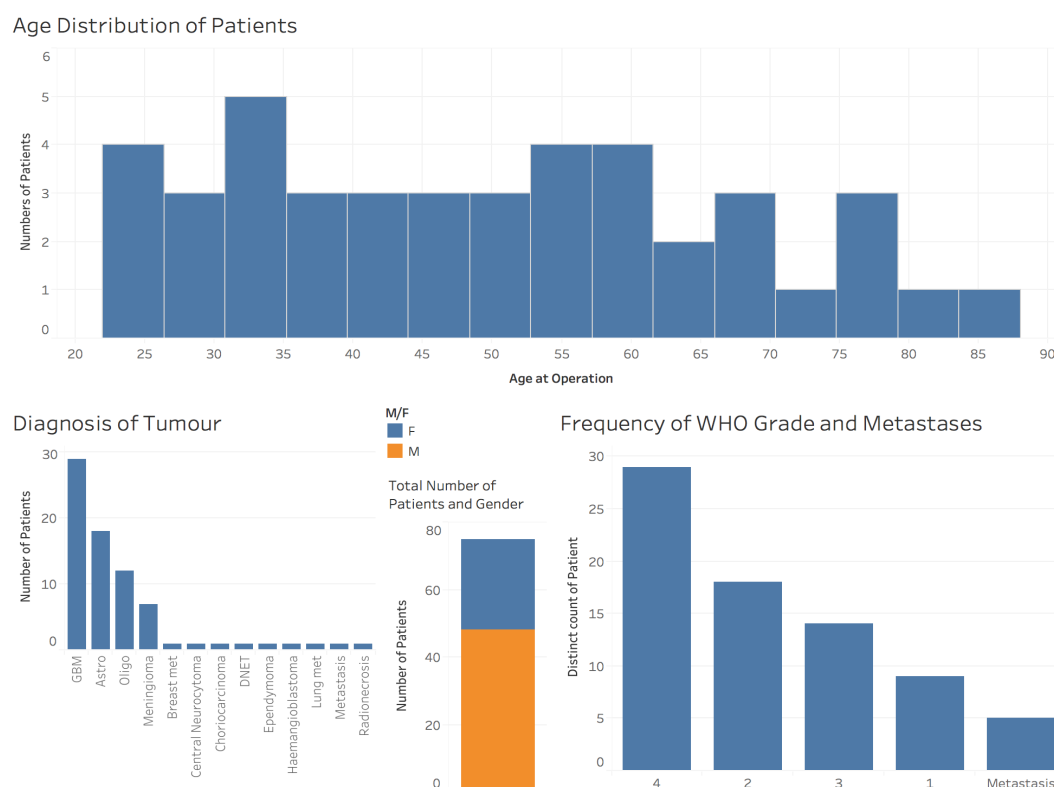


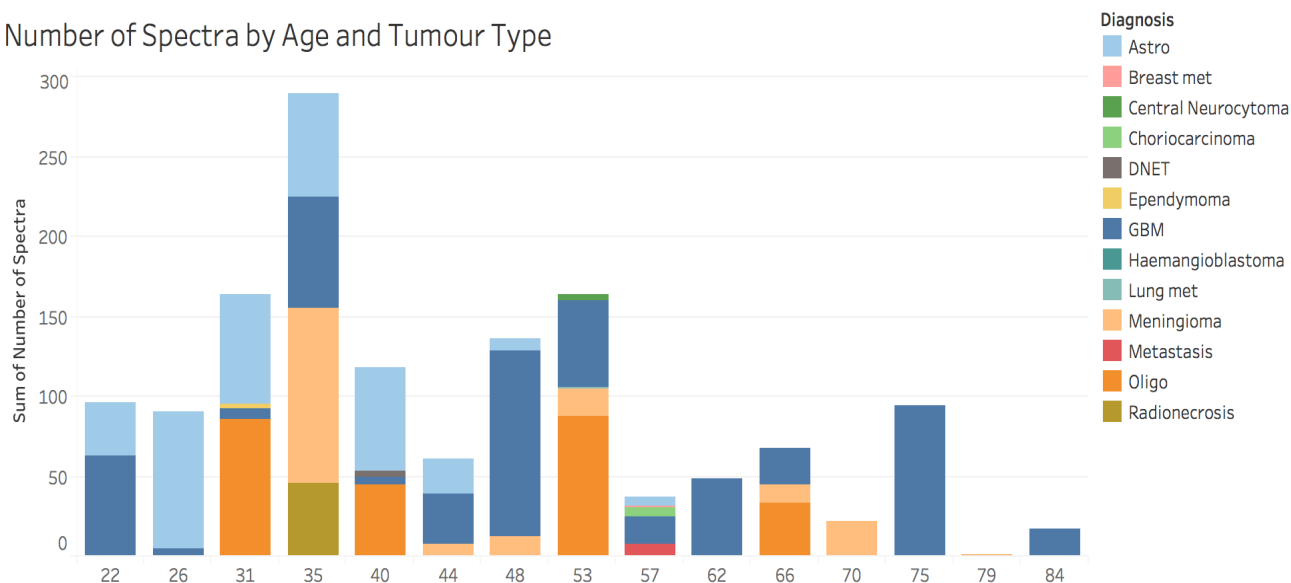
Figure 124. Demographic summary of all patients in the iKnife trial
Age (top), gender (bottom centre) and pathology (bottom left and right) characteristics of patients included in the iKnife study.

A total of 27 females and 48 males were recruited. The ages of patients recruited ranged from 22 to 83 years old with a mean age of 48. A range of common brain tumour diagnoses were represented including glioma and non-glioma tumours, together with all WHO grades as shown in the summary Figure 124.

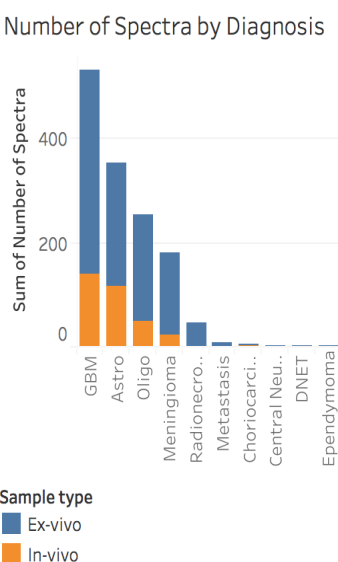
I collected a total of 1388 spectra. Caution should be exercised in interpreting the number of spectra as a measure of data quantity. Raman is non-averaged point measurement in time and space but REIMS data is a recording of molecules entering the mass spectrometry machine averaged over time. One spectrum can last many seconds and can therefore contain a lot of data. Numbers of spectra are still used here as a rough gauge of the amount of data collected as discrete units. Figure 125 below shows some important trends that I have observed.

1. There is far more *ex vivo* than *in vivo* data
2. More males than females
3. WHO grades 4 and 2 most common, with grades 3 and 1 following after
4. As expected, tumours have an age pattern with lower grade gliomas being more common in younger patients, and GBMs being more common in elderly patients.

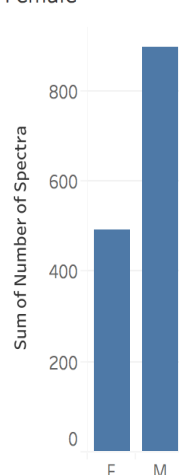
Number of Spectra by Age and Tumour Type



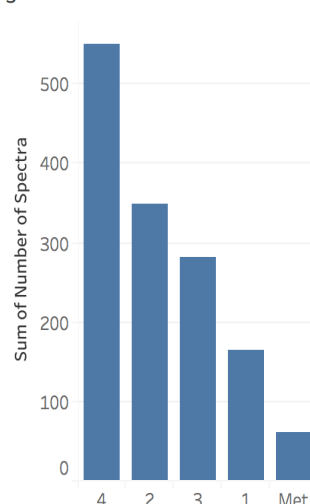
Number of Spectra by Diagnosis



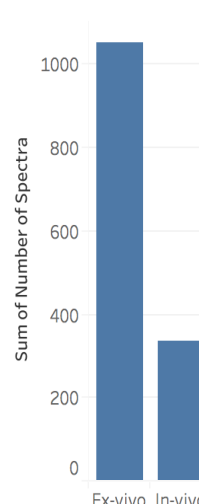
Number of spectra: Male vs Female



Number of Spectra by WHO grade



Number of spectra: in vivo vs ex vivo



Number of Tissue Samples Reported in vivo and ex vivo

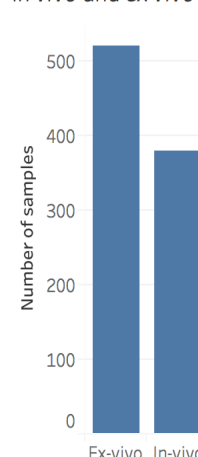


Figure 125. Detailed breakdown of the demographic characteristics of the spectra collected.

Top: Age and tumour type, Bottom (left to right): Diagnosis, male vs female, WHO grade, in vivo vs ex vivo and number of tissue samples.

I performed unsupervised multivariate analysis on male/female spectral data. The PCA plot in Figure 126 shows no obvious separation but there is some clustering (as would be expected).

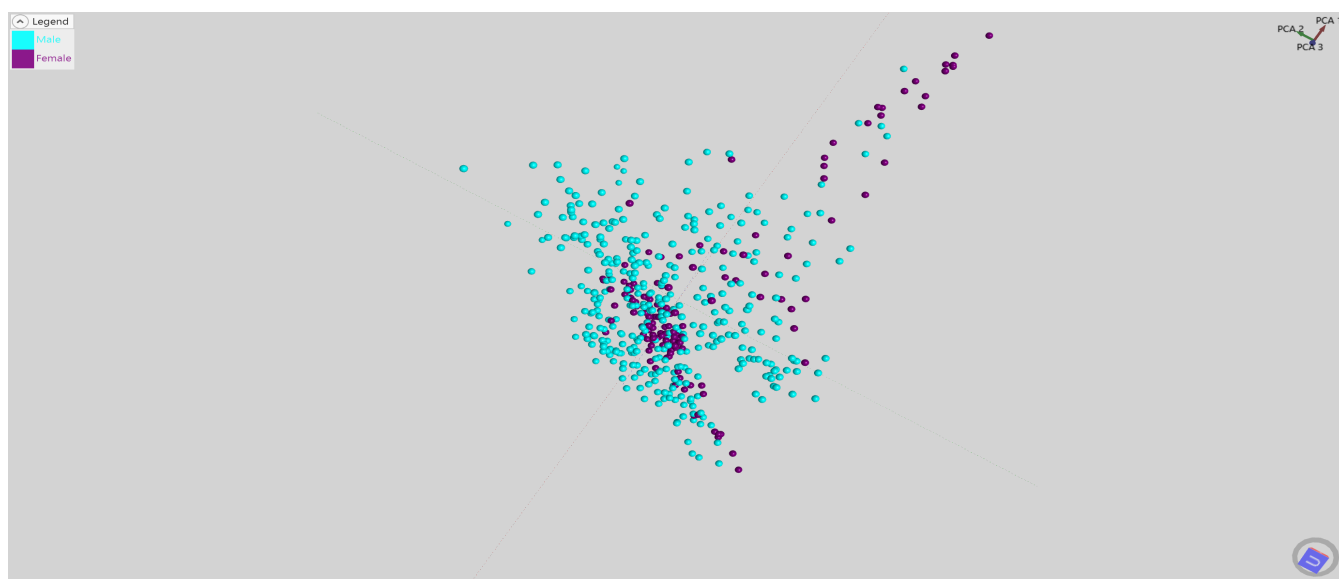


Figure 126. PCA plot of the top 3 principle components comparing Gender of patients in the study.

No clear separation is seen between brain spectra obtained from male patients and female patients.

LDA does show male/female difference more clearly in Figure 127. This would be expected as there are significant genetic and as a result molecular differences between males and females.

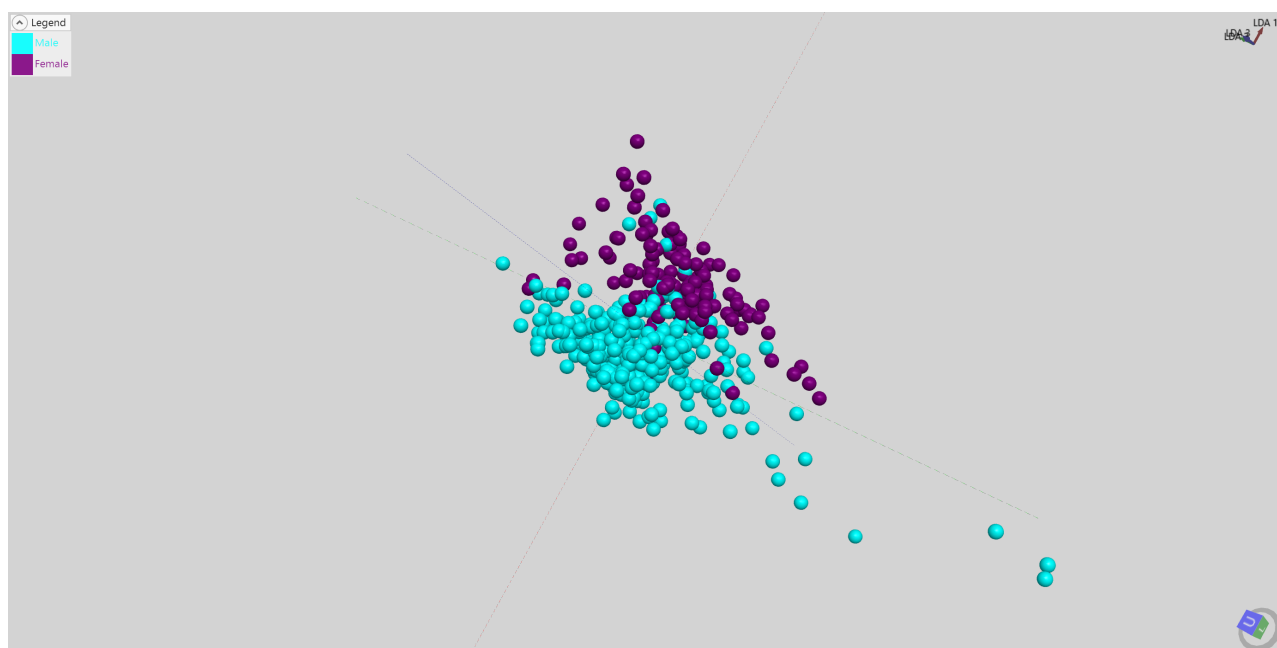


Figure 127. LDA plot of the top 3 LDA components comparing Gender of patients in the study.

Male and female spectra cluster clearly on LDA.

Patient age does not cluster on PCA plot as shown below in Figure 128.

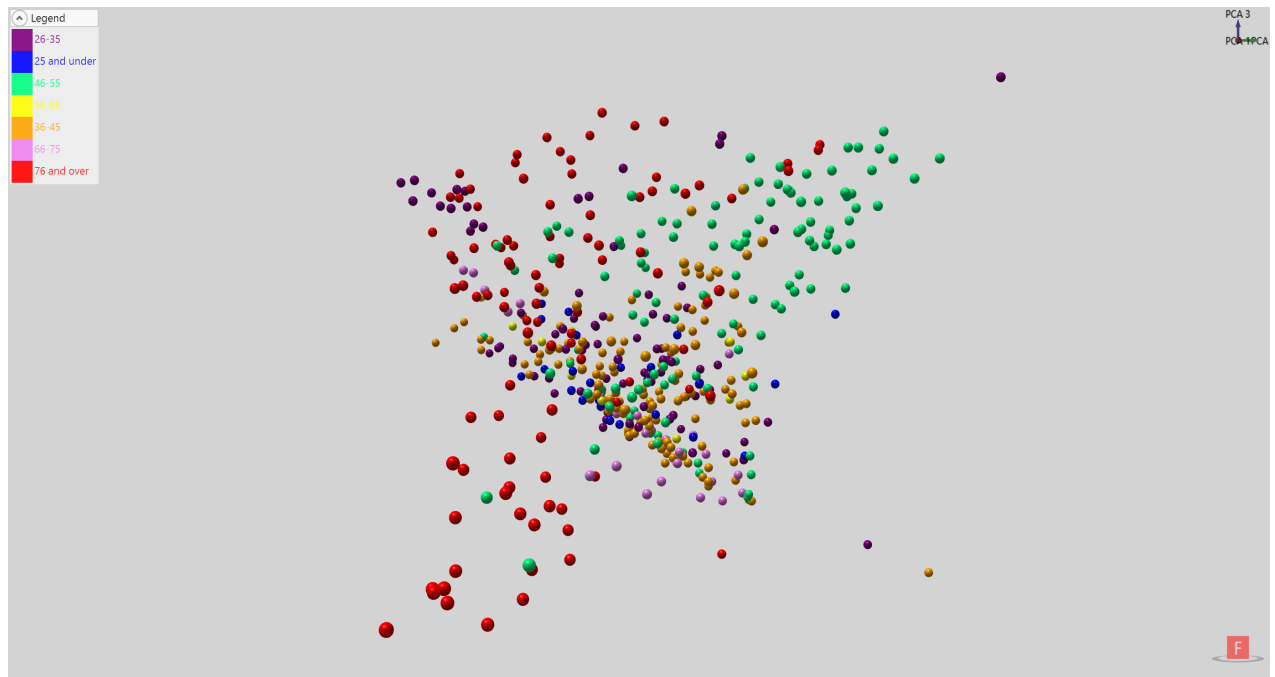


Figure 128. PCA plot of the top 3 principle components comparing the age of patients in the study
No clear separation can be seen.

Patient age does show some clustering on LDA as shown in Figure 129. There is some separation between age 46-55 and 76 and over, otherwise separation is not clear.

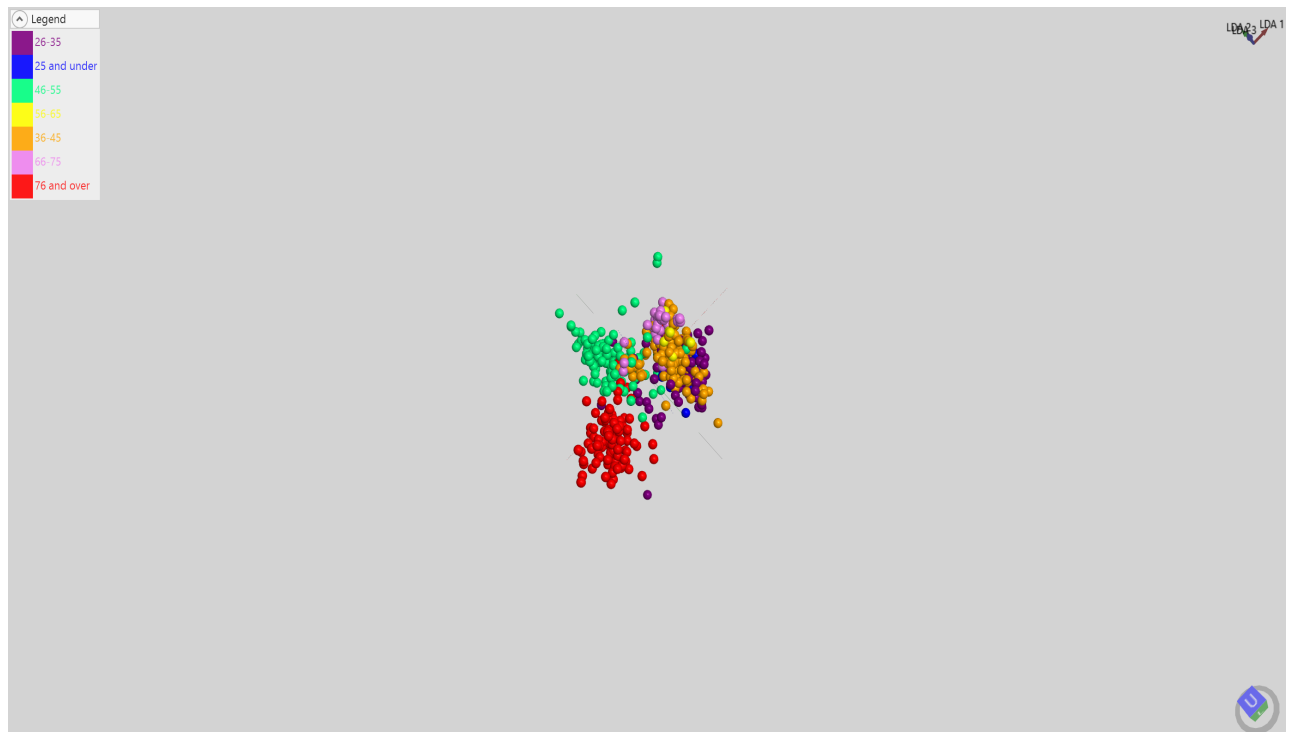


Figure 129. LDA plot of the top 3 LDA components comparing the age of patients in the study.
Some age groups cluster well on LDA, particularly 76 and over and 46-55 age groups.

5.5.2.3. Summary

I found that patient age and gender do not separate on PCA, with some separation on LDA. The distribution of age and gender shows no obvious bias (slightly more males than females).

5.5.3. *In vivo* vs *Ex vivo*

5.5.3.1. Introduction

The main question to answer is: how similar are the *in vivo* and *ex vivo* datasets? This has implications on whether data from *ex vivo* tissue can be used to create or augment an *in vivo* model.

The hypothesis is that the molecular nature of non-fixed *ex vivo* tissue will be similar to *in vivo* tissue as tissue fixation methods will not have been used and the lipidomic data will be preserved. There will therefore be a negligible difference between the spectra.

5.5.3.2. Multivariate analysis

The REIMS system was used to collect data during surgery (*in vivo*) and on samples taken after surgery (*ex vivo*). *Ex vivo* data forms the largest of the groups of data collected, therefore one main question was if there is a difference between *in vivo* and *ex vivo* data. This has important implications for database building. Our hypothesis was that the data will be similar as the phospholipids will be preserved in frozen tissue. LDA analysis reveals, as predicted, no clear separation: *in vivo* is teal, *ex vivo* is blue on plot in Figure 130.

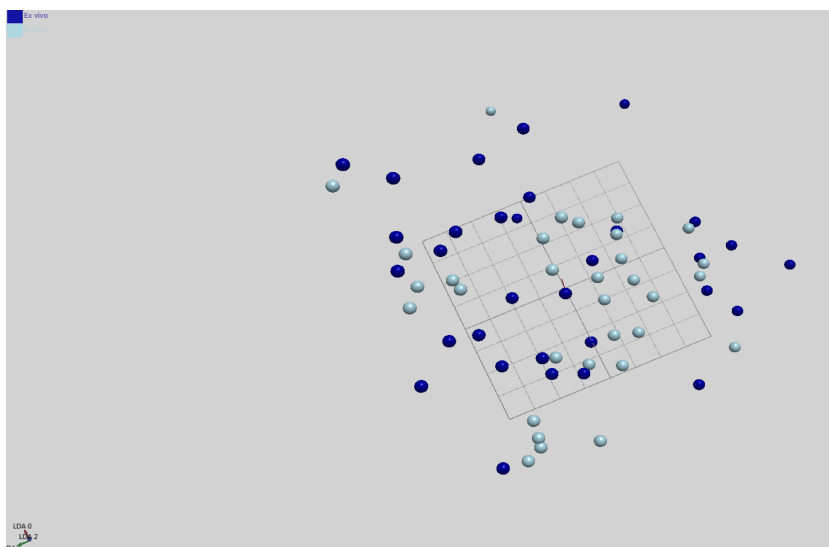


Figure 130. 3-D representation of PCA of *in vivo* vs *ex vivo* iKnife spectra. LDA analysis of *in vivo* vs *ex vivo* datasets. There is no clear separation between both groups.

Cross validation results confirmed a poor ability to differentiate *in vivo* from *ex vivo* data (LOOCV=33%).

5.5.3.3. Conclusion

There are no strong contraindications to combining *in vivo* and *ex vivo* datasets for analysis. This increases the amount of data available for model building by allowing *ex vivo* data to contribute to the classification model for *in vivo* use. All following models have combined *ex vivo* and *in vivo* data.

5.5.4. IPA vs Non-IPA

5.5.4.1. Introduction

All IPA cases were compared with non-IPA cases to see if a difference could be observed using PCA, LDA and cross validation. This was important in order to determine if data from IPA and non-IPA cases can be combined in the same model. IP and non-IPA spectra were labelled and loaded into OMB. PCA/LDA models and cross validation was performed to see if the 2 datasets were detected as different.

5.5.4.2. Multivariate analysis

A total of 47 tissue samples and 90 spectra were analysed. 43 IPA and 47 non-IPA spectra were compared by loading into OMB, selecting appropriate tissue burns. One spectrum per burn was averaged, lockmass being used for IPA spectra of 544.2615. Figure 131 shows the resultant PCA plot.

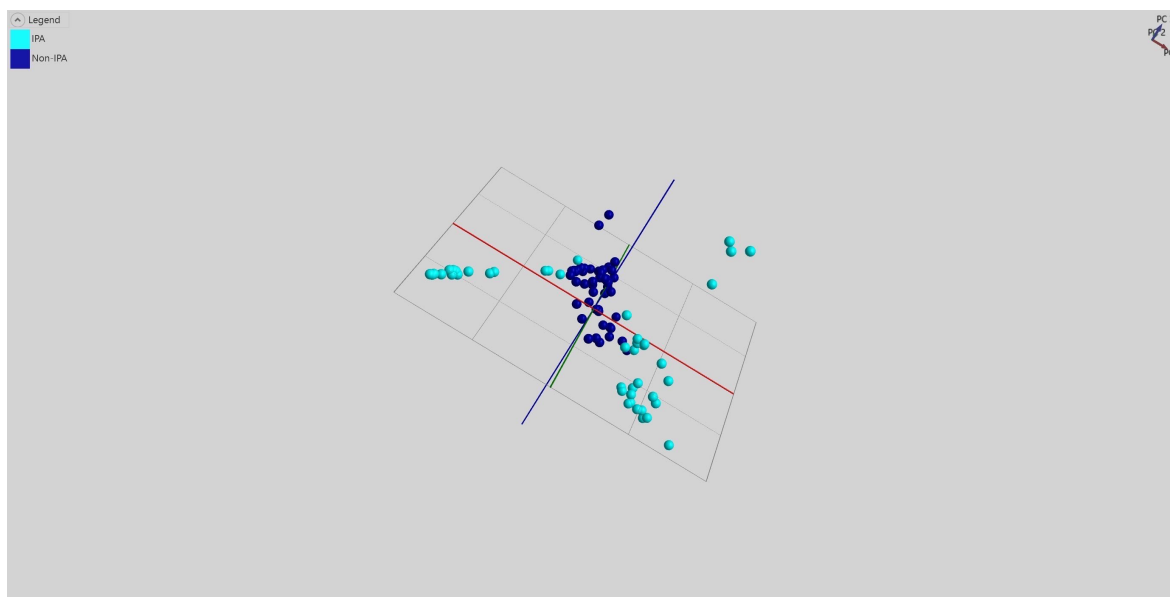


Figure 131. PCA analysis of IPA vs non-IPA spectra.

Clear differences are demonstrated with tight clustering of non-IPA spectra.

IPA and non-IPA spectra separate well on PCA. LOPOCV was performed on the PCA model, shown in Table 22 with a result of 94.4%. The model can therefore accurately distinguish IPA from non-IPA spectra.

Confusion matrix:

	ipa	non-ipa	Total
ipa	39	4	43
non-ipa	1	46	47
Total	40	50	90

Table 22. Confusion matrix from cross validation of IPA vs non-IPA using a PCA model, m/z 100-1000.

5.5.4.3. Conclusion

IPA spectra and non-IPA spectra have been demonstrated as significantly different. Using multivariate modelling the system can distinguish between IPA and non-IPA spectra. Combining the data from IPA and non-IPA methods for comparisons may therefore not be feasible. Models will therefore be presented as IPA and non-IPA models. In some cases, the models were combined out of interest in case the differences between the classes being compared were greater than the differences between IPA and non-IPA spectra.

5.5.5. Tumour Vs Normal

5.5.5.1. Introduction

In the previous results chapters the spectra, samples and patient characteristics of the core dataset have been established. It has been determined that predictive models cannot be trained to separate *in vivo* and *ex vivo* data using unsupervised or supervised machine learning methods. These data groups will therefore be combined for the following analyses to ensure the training and testing data sets are as large as possible for model building. IPA and non-IPA have been shown to be datasets which are significantly different enough to interfere with model building for other features. This data will therefore be analysed separately.

The most important clinical question is whether REIMS data when used *in vivo* can distinguish between tumour and non-tumour tissue. The results of this analysis will be presented here. The hypothesis is that REIMS data can be used to create models able to distinguish tumour from non-tumour tissue based on the molecular differences measured in the mass spectrum.

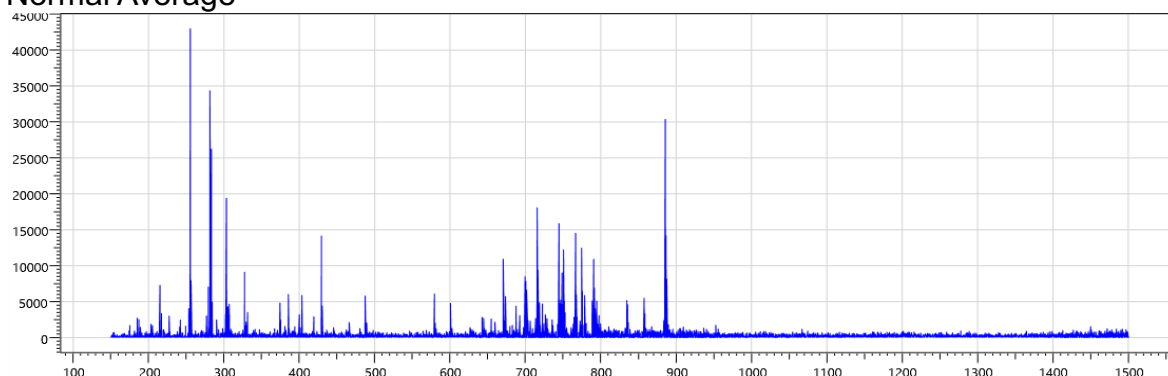
Non-IPA, IPA and then a combined IPA and non-IPA analysis will be presented.

5.5.5.2. Average analysis

Over 991 spectra were collected in the iKnife study, yielding 639 tumour and 352 normal spectra. 146 normal and 309 tumour samples were collected for analysis from 75 patients. The averaged spectra are presented in Figure 132.

Average Spectra

Normal Average



Tumour Average

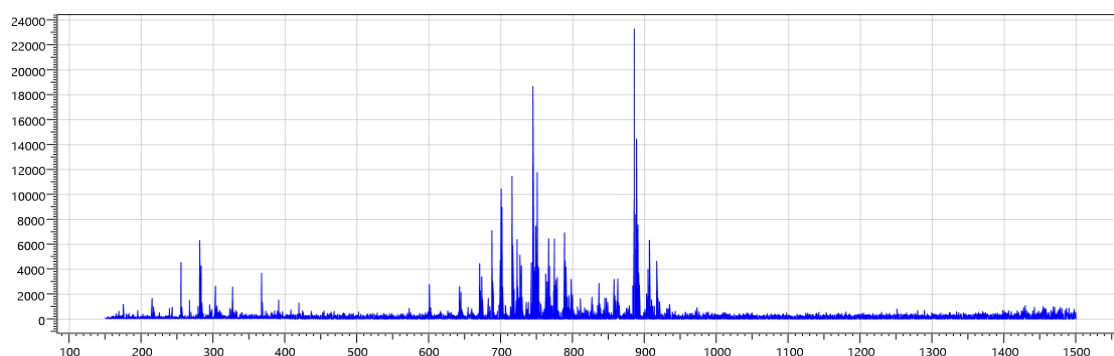


Figure 132. Averaged mass spectra for normal and tumour tissue.

Top: normal tissue spectrum, bottom: average tumour spectrum. X-axis: m/z ratio, Y-axis: Relative intensity.

The clear difference is that tumour samples contain more signal in the higher phospholipid mass ranges, whereas normal samples contain more signal in lower (triglyceride) mass ranges. This forms a good basis for inferring that good performance models can be built to classify tumour versus normal tissue.

5.5.5.3. Multivariate analysis

For multivariate analysis the spectra were grouped into non-IPA spectra and IPA spectra. Non-IPA results will be presented first.

Spectra noted as 'Normal' by histopathology were compared to a random selection of tumour spectra, by direct infusion (m/z 100-1000). Spectra was averaged for all burns due to variability in signal. PCA analysis shows clustering of normal and tumour samples in Figure 133. This suggests similarities exist within tumour and normal classes and differences exist between tumour and normal to allow separate clusters to form on a 3D PCA plot.

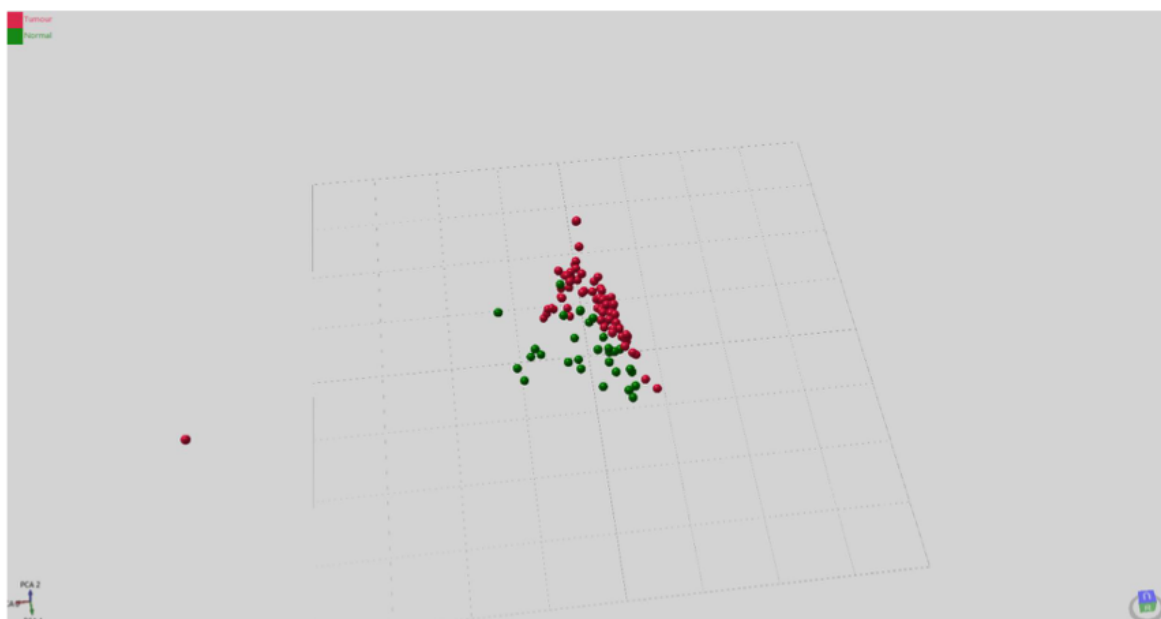


Figure 133. PCA analysis of tumour vs normal samples, non-IPA, m/z100-1000. *There is clear separation between tumour (red) and normal (green) spectra using the first 3 principle components.*

LDA analysis of matched tumour vs normal samples confirms strong clustering of tumour and normal samples as shown in Figure 134.

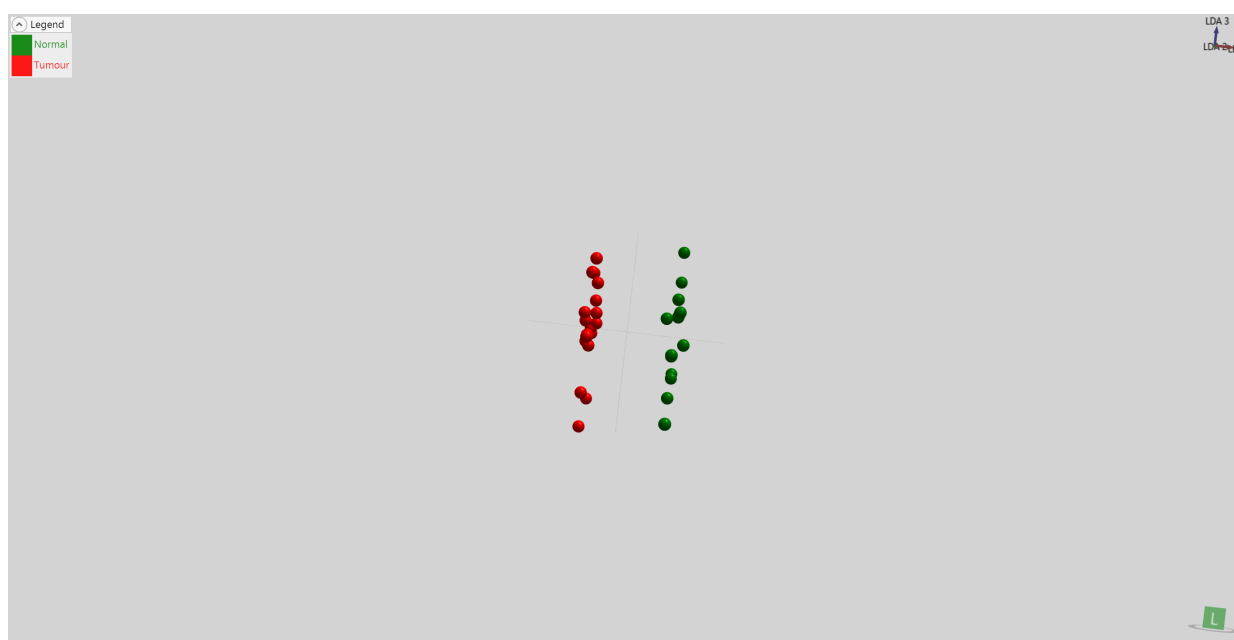


Figure 134. LDA analysis of tumour vs normal samples, non-IPA, m/z100-1000. *Clear separation between groups is shown. Tumour (red) and normal (green) matched samples were used for this plot.*

Cross validation using Leave One Patient Out (LOPOCV) on non-IPA data from the PCA-LDA model is shown in the cross validation matrix in Table 23.

Cross-validation matrix:

	Normal	Tumour	Total
Normal	11	3	14
Tumour	2	16	18

Table 23. Confusion matrix of the LOPOCV results for IPA vs Non-IPA comparison, PCA-LDA model, m/z 100-1000.

The model shows a clear ability to differentiate between normal and tumour spectra.

Using this cross validation matrix an accuracy of 84.3% is reached:

Sensitivity= $11/14=79\%$

Specificity= $16/18=89\%$

Accuracy= $11+16/14+18=84.3\%$

Using a narrower m/z 600-1000 yields similar accuracy on cross validation as shown in Table 24.

Cross-validation report - Brain - tumour v normal no IPA 600-1000 20161222

Model name Brain - tumour v normal no IPA 600-1000 20161222
Creation time 22/12/2016 09:36:25

All out at once No
Groups to leave out 13,17,19,23,3,31,8,9
Group percentage out 100
Outlier type None

Group result:

Group	Number of spectra	Number of passes	Number of failures	Number of outliers	Correct Classification Rate
13	2	2	0	0	100%
17	6	3	3	0	50%
19	3	3	0	0	100%
23	13	10	3	0	76.92308%
3	2	1	1	0	50%
31	8	8	0	0	100%
8	4	4	0	0	100%
9	9	8	1	0	88.88889%
Total	47	39	8	0	82.97872%

Confusion matrix:

	normal	tumour	Total
normal	17	2	19
tumour	6	22	28
Total	23	24	47

Table 24. Confusion matrix of the LOPOCV results for the non-IPA PCA/LDA model, m/z 600-1000.

The model shows a clear ability to differentiate between normal and tumour spectra with only a marginal difference with the m/z 100-1000 model.

The loadings for this PCA-LDA model calculated using OMB in Figure 135 show the crucial data that separates normal from tumour spectra lies in the higher mass ranges, with a particularly strong peak near m/z 750.

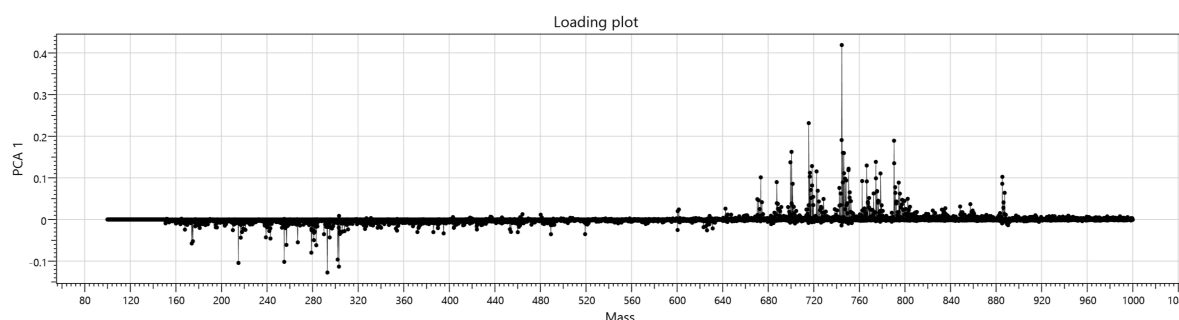


Figure 135. Non-IPA PCA-LDA model loadings for normal vs tumour.

The loadings based on the PCA plot show that the segregation observed is due to higher mass ranges, as suggested crudely by the normal and tumour averaged spectra.

IPA samples represent a larger dataset for analysis as IPA was used for the majority of the patients in the study. 447 spectra were analysed from 97 tissue samples categorised as Normal and Tumour as defined by the histopathological report. Spectra were excluded only if SNR was unacceptable as defined previously (see iKnife method development in section 4.2). The IPA analysis involved 35 patients. Lockmass was set at 554.2615. One spectrum was recorded per burn. The LDA analysis showed clear separation of tumour from normal as shown in Figure 136.

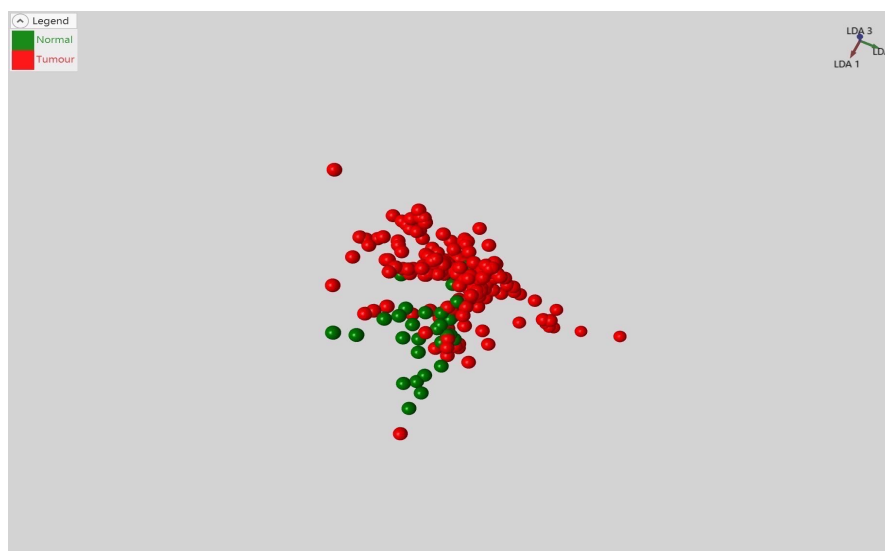


Figure 136. LDA plot of tumour vs normal for IPA spectra, m/z 100-1000.

Clear separation between normal (green) and tumour (red) spectra is seen.

Cross Validation showed an accuracy of 86.3% in distinguishing tumour from non-tumour as shown in Table 25.

	Normal	Tumour	Outlier	Total
Normal	23	7	1	30
Tumour	20	147	17	167

Table 25. Confusion matrix of leave 20% out cross validation of the PCA-LDA model of IPA data for tumour vs normal.

154 tumour and 43 normal spectra were tested using cross validation.

Using this model, the following classification performance statistics were calculated:

86% of tumour samples successfully classified, 77% of normal samples successfully classified.

The sensitivity and specificity were calculated as:

Sensitivity=88%

Specificity=77%

Accuracy=86.3%

Lastly, I combined IPA and non-IPA data to see if the predicted poor performance in section 4.5.4 was indeed seen in practice. Given human brain tissue is such a rare commodity it would be wise to exhaust any options to combine data. 991 spectra were analysed (one spectrum per burn), 304 normal and 686 tumour, from 146 normal and 309 tumour samples. All 75 patients were included in this model. The mass range used was 100-1000 (widest range possible). The large number of patients in this analysis allows the use of LOPOCV without removing too much data each time the model is built.

The scores and loadings plots from the PCA-LDA model are shown in Figure 137. These show reasonable separation based on Principle Components 1 and 2. The loadings plot shows higher mass ranges to be more responsible for the separation.

Score and Loading Plots

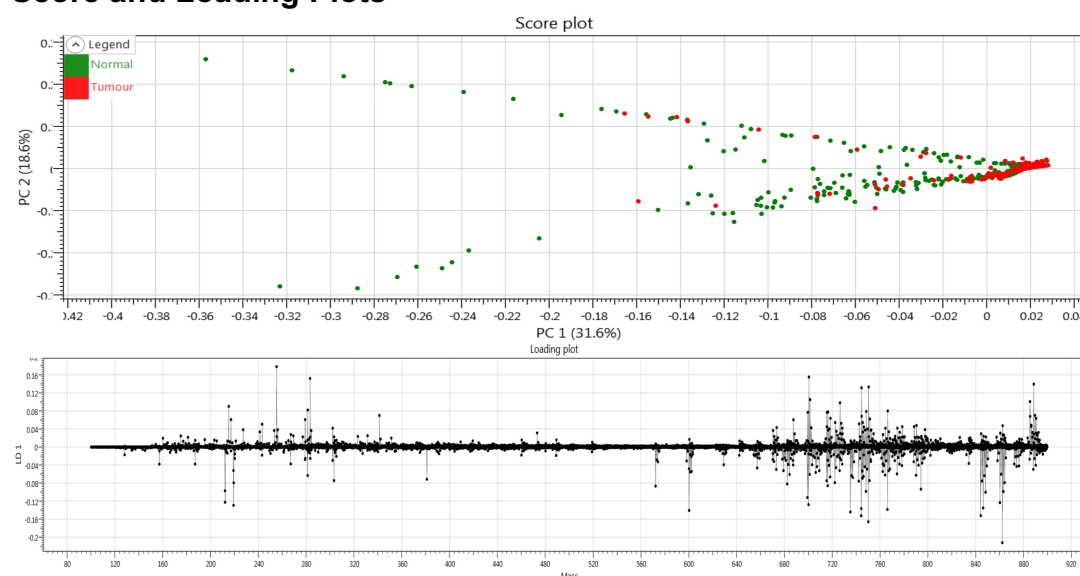


Figure 137. Scores and loadings from the combined IPA-nonIPA PCA-LDA model.

Scores (top) and loadings (bottom) based on 2 principle components show low and high m/z values contribute to the model.

The confusion matrix In Table 26 shows reasonable performance of this classifier.

Confusion matrix:

	normal	tumour	Outlier	Total
normal	240	110	2	352
tumour	82	556	1	639
Total	322	666	3	991

Table 26. LOPOCV confusion matrix of tumour vs normal tissue, IPA and non-IPA data combined, m/z 100-1000, lockmass 554.2615.

Overall my accuracy using LOPOCV for non-IPA and IPA spectra combined is 80.1%.

5.5.5.4. Conclusion

The iKnife is able to distinguish normal from tumour. The accuracy can be as high as 86.3% using IPA in this current dataset. With a larger data series and more normal tissue it is likely the accuracy will increase further. Combining IPA and non-IPA data results in a lower accuracy of 80%. The system can successfully distinguish tumour from non-tumour as a proof of concept, with higher accuracy using IPA.

5.5.6. Tumour Density

5.5.6.1. Introduction

Hypothesis: Due to its high dimensionality REIMS should be able to detect the quantity of tumour cells in brain tissue.

REIMS spectra for tumour density were investigated by classifying spectra according to tissue density of tumour cells (as reported by the histopathologist). The categories were graded from 0% tumour to 100% tumour with necrosis:

Normal: None
 Tumour: Low
 Tumour: Moderate
 Tumour: High
 Tumour plus necrosis

I created models with all IPA and non IPA spectra, IPA only data and non-IPA only

data and with a variety of mass ranges: 100-1000, 600-900 and 600-1000. These ranges were based on tumour density models from breast and ovarian cancer. The best performing models are presented here.

5.5.6.2. Multivariate analysis

64 spectra from 35 patients were analysed comprising 35 spectra from high density and 29 spectra reported as no cancer cells. PCA-LDA models were created using non-IPA and IPA spectra. The PCA plot showed a very strong clustering effect between these tumour density categories in Figure 138.

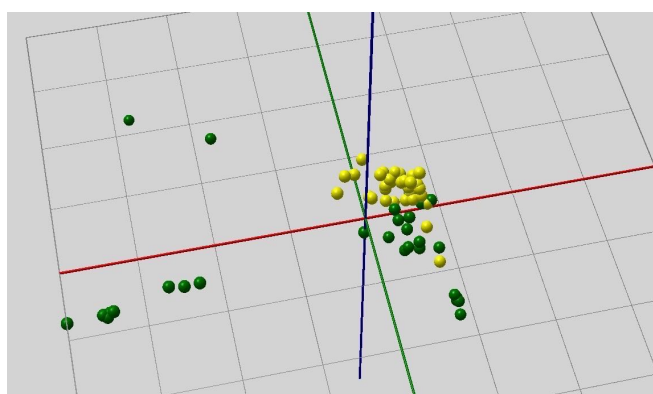


Figure 138. PCA plot of tumour cell density, none vs high, non-IPA spectra, m/z 100-1000.

Good separation is seen between low (green) and high (yellow) density tumour spectra.

Assessing the associated PCA-LDA model using 20% out cross validation confirms an excellent accuracy of 89.1% for non-IPA data as shown in Table 27.

Confusion matrix:

	high	none	Total
high	34	1	35
none	6	23	29
Total	40	24	64

Table 27. Leave 20% out cross validation of the tumour density non-IPA model.

IPA spectra were then analysed. Data from 15 patients included 141 spectra with 37 high and 104 moderate density labelled spectra. The data was lockmassed, background subtracted and normalised with an m/z range 100-1000. High and moderate tumour density IPA spectra separate very well on PCA as shown in Figure 136 with hardly any overlap. This predicts a strong classifier model.

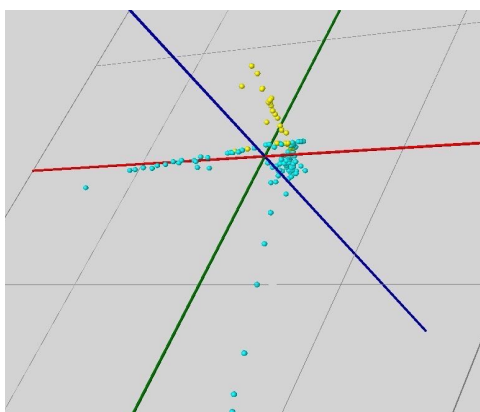


Figure 139. PCA plot of moderate vs high density tumour tissue spectra. IPA, m/z100-1000.

The PCA results in Figure 139 are predictive of good cross validation, in this case using LOPOCV an impressive overall accuracy of 88.5% is achieved as shown in the confusion matrix in Table 28.

Confusion matrix:

	high	mod	Outlier	Total
high	33	3	1	37
mod	13	90	1	104
Total	46	93	2	141

Table 28. Confusion matrix from LOPOCV of the moderate vs high tumour density IPA model.

In order to assess finer classification of tumour density 4 classes were created using 320 tissue samples from 39 patients. These consisted of 639 spectra classified as normal, tumour low, tumour moderate, tumour high and tumour necrosis. IPA and non-IPA data was combined and models were built in the 100-1000 mass range. When combining all possible classes of tumour density LDA showed good segregation according to tumour density as shown in Figure 140.

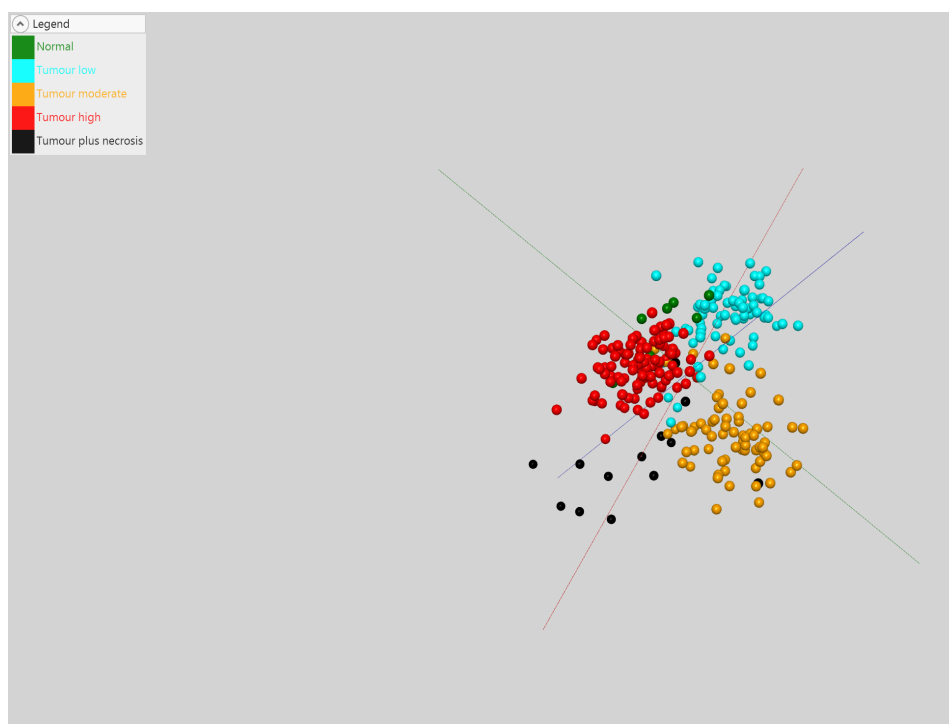


Figure 140. LDA plot of tumour density using IPA and non-IPA spectra, m/z 600-800.

Excellent segregation of tissue samples based on tumour cell density as reported by histology is seen. Necrotic tumour areas also segregate well.

Cross validation using the PCA-LDA model for 4 density classes was performed with LOPOCV and returned an accuracy result of 59%.

5.5.6.3. Conclusion

I found that splitting the groups into 2 simple categories increases the classifier accuracy significantly:

- Normal vs High 89% accuracy with excellent PCA (IPA and non-IPA)
- Mod vs High 88.5% accuracy using LOPOCV (IPA)

Very poor performance was seen when multiple categories are the output of the algorithm despite good LDA plots separating density classes well. This may mean more data is required to power algorithms with more than 2 tumour density classes.

5.5.7. Tumour Type

5.5.7.1. Introduction

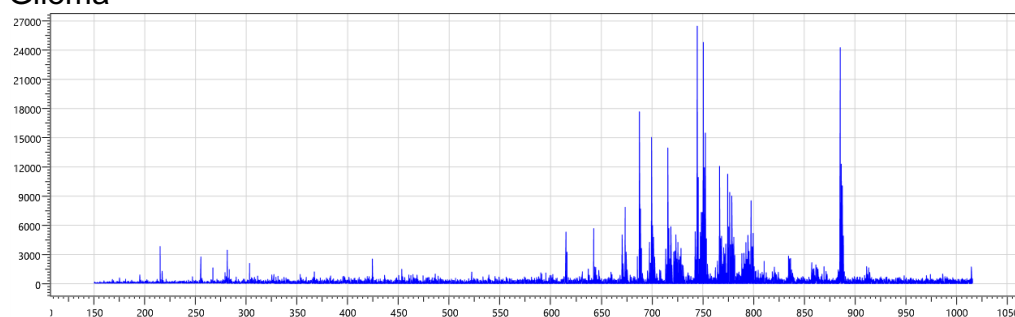
Several comparisons were attempted. The most basic difference between samples was whether they were from glioma or non-glioma tissue. The next analysis performed was with regards to the specific histological identity of the tumour: this was more than just a simple binary classification with 6 different diagnostic classes being included. The hypothesis was that iKnife data can be used to create models able to distinguish different tumour types based on molecular differences between tumours.

5.5.7.2. Average analysis

87 spectra from 86 samples were analysed consisting of 67 Glioma and 20 Non-Glioma samples from 16 patients matched for age and gender without IPA. Due to variation in intensity one spectrum per sample was analysed with lockmass correction, background subtraction and normalisation as described in section 4.4.3.

Simple analysis of the average spectra from these 2 groups shows a clear difference. Glial spectra are rich in data in the higher mass range. Non-glioma spectra contain more data in lower mass ranges and have fewer peaks in the higher mass range. The average spectra are shown in Figure 141.

Glioma



Non Glioma

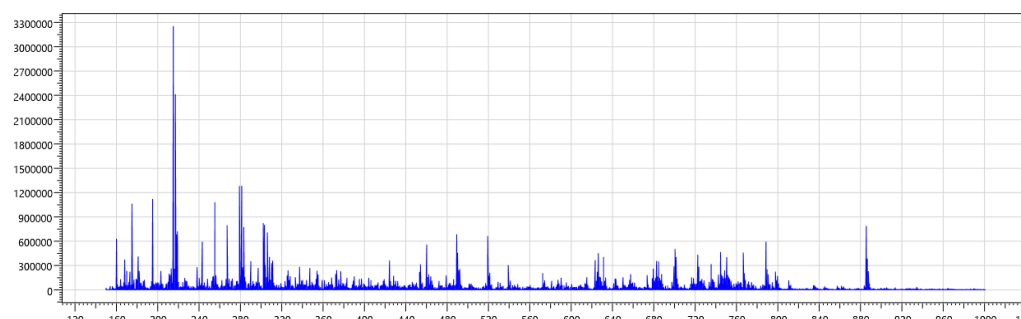


Figure 141. Average mass spectra, m/z range 100-1000, for glioma (top) and non-glioma tissue (bottom), non-IPA.

Glioma and non-glioma spectra differ in that most data is contained in the high m/z range for glioma and the low m/z range for non-glioma.

The average spectra from different tumours are visually different, as shown in Figure 142. Key peaks have been labelled. This would suggest that specific models can be created for each tumour type.

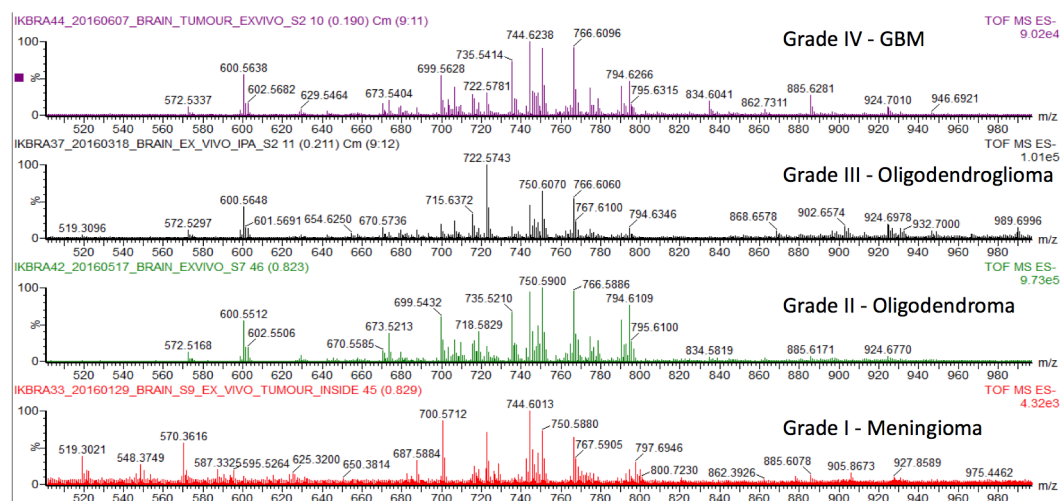


Figure 142. Comparison of the average spectrum of common brain tumours using REIMS, m/z 500-1000.

Differences throughout the whole m/z spectrum are seen for multiple tumour types spanning different tumour grades based on the WHO scale.

The difference between glioma and non-glioma average spectra is more striking than within the different tumour types.

5.5.7.3. Multivariate analysis

Unsupervised analysis using PCA was performed revealing separation between glioma and non-glioma groups with spectra matched for age and gender as shown in Figure 143.

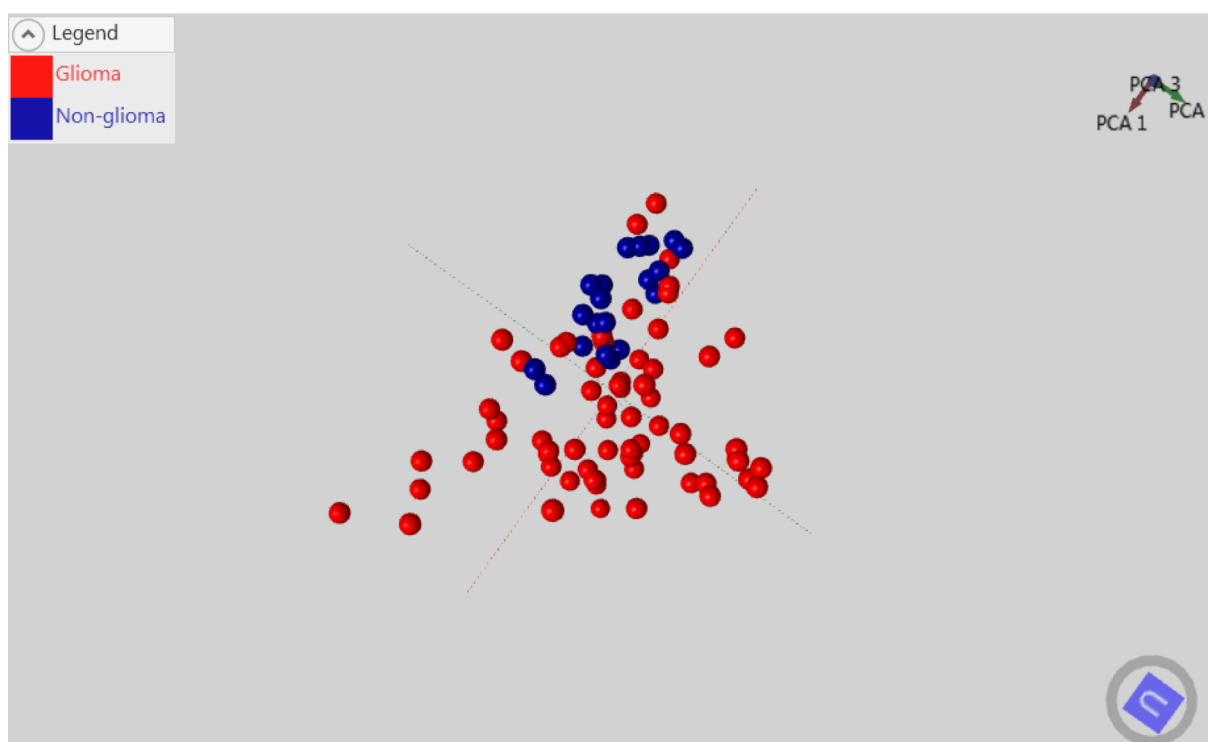


Figure 143. PCA analysis of glioma vs non-glioma in 16 selected patients matched for age and gender, m/z600-1000, non-IPA.

Glioma spectra are red coloured spheres and non-glioma are blue spheres.

Supervised analysis using LDA revealed excellent separation between the 2 classes with hardly any overlap as shown in Figure 144.

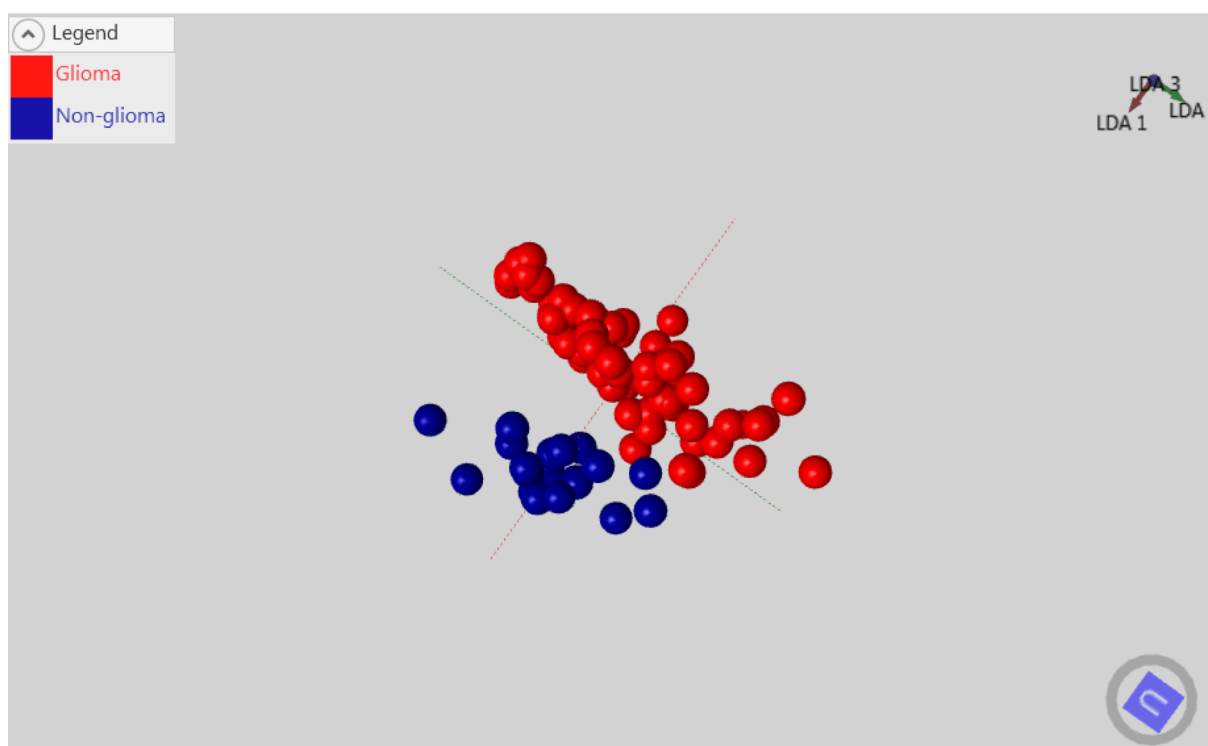


Figure 144. LDA analysis of glioma vs non-glioma in 16 selected patients matched for age and gender, m/z600-1000, non-IPA.

Excellent separation between glioma spectra (red spheres) and non-glioma spectra

(blue spheres)

Cross validation using LOPOCV was performed on the PCA-LDA non-IPA model with the results shown in Table 29.

Cross-validation matrix:

	Glioma	Non-glioma	Total
Glioma	63	3	66
Non-glioma	1	19	20

Table 29. LOPOCV results for non-IPA REIMS data.

The cross validation matrix shows excellent performance of this model with 95.34% accuracy achieved.

IPA data using leave one out cross validation gave the cross validation matrix in Table 30 with an accuracy of only 70 %

Confusion matrix:

	glioma	non-glioma	Outlier	Total
glioma	220	84	23	327
non-glioma	44	53	10	107
Total	264	137	33	434

Table 30. LOPOCV results for IPA REIMS data.

The basis for the good PCA is separation using the highest mass ranges- in particular 880-900 range as shown in the loadings plot in Figure 145.

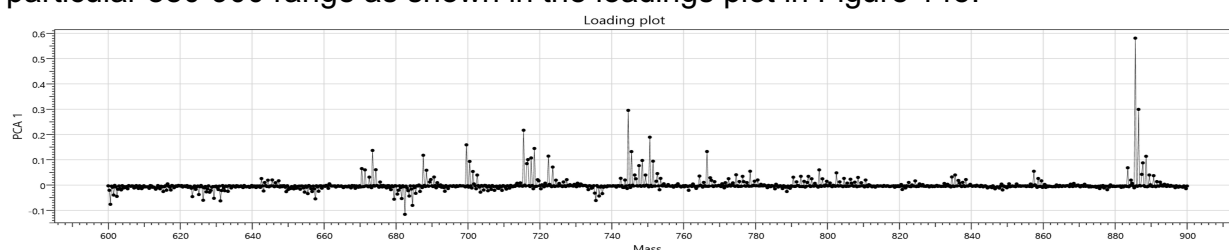


Figure 145. Loadings plot of the PCA-LDA model for non-IPA glioma vs non-glioma REIMS model.

Multivariate analysis using PCA-LDA modelling was performed on IPA and non-IPA data for histological classes. The PCA plot in Figure 147 shows reasonable separation between tumour types.

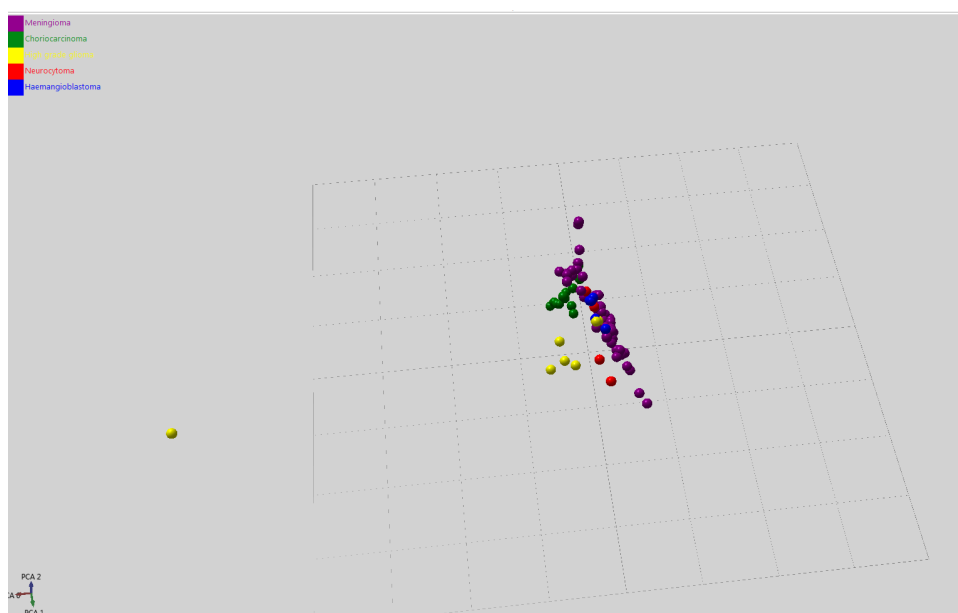


Figure 147. PCA analysis of tumour type histology, m/z600-1000, non-IPA.

The LDA model from IPA data shows good clustering between the 6 tumour types as shown in Figure 148.

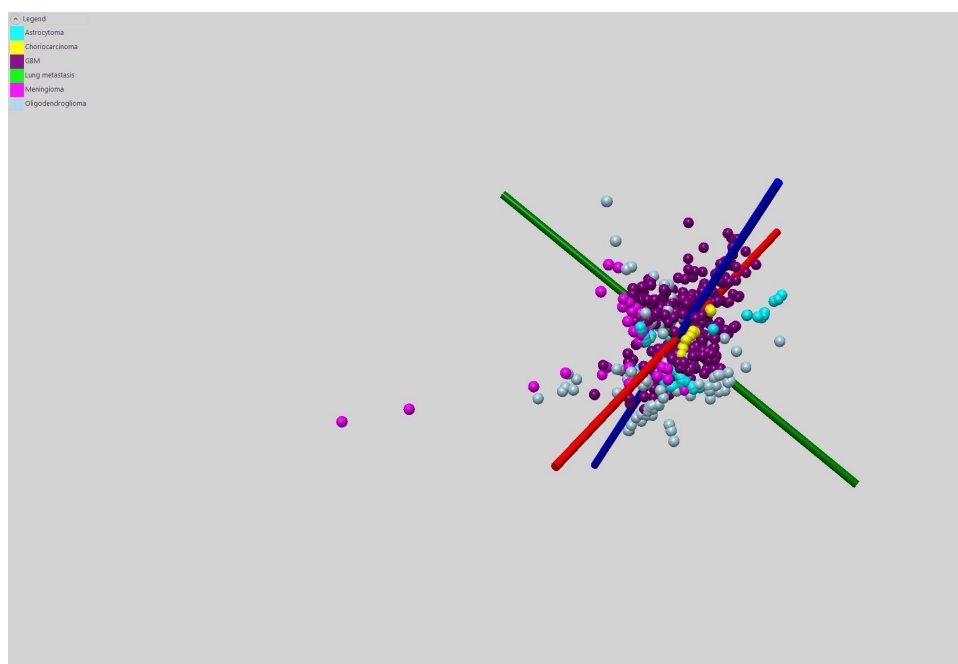


Figure 148. LDA analysis of tumour type histology, m/z600-1000, non-IPA.

There is good separation of the 6 histology types in this study. The 3 types of glioma separate very well.

Cross validation results of the PCA-LDA model in Table 31 reveals a low accuracy of 47.1% with leave 20% out cross validation.

Model	LOPOCV	Leave 20% Out CV
IPA	29.3	47.1
Non-IPA	25.0	42.6
Combined	19.7	33.1

Table 31. Cross validation performance of PCA-LDA models using IPA, non IPA and combined data.

5.5.7.4. Conclusion

The best performing model for identifying tumour type is glioma vs non-glioma without IPA, achieving an impressive cross validation accuracy of 94%. Diagnosing the exact tumour is more difficult. With 6 classes, IPA and non-IPA data being incompatible for model construction means not enough data is available to train the good PCA-LDA model suggested by the LDA results.

5.5.8. Tumour Grade

5.5.8.1. Introduction

Model building was attempted with tumour grade as the identity class. 2 models were investigated. The first model involved classifying tumours into low grade (WHO grades I and II) and high grade (WHO grades III and IV). The second model classified tumours into all 4 WHO grades. WHO grading is important to determine as it relates directly with treatment choices, with the main groups being low grade and high grade disease.

The hypothesis was that REIMS spectra may separate different grades of tumour based on the clinic-pathological basis of tumour grading which relates in some way to tumour biology.

5.5.8.2. Multivariate analysis

My PCA analysis of IPA data showed good separation of low grade and high grade spectra. Data was from 35 patients consisting of 480 spectra from 191 samples, 123 low grade spectra and 357 high grade spectra. The PCA plot is shown in Figure 149.

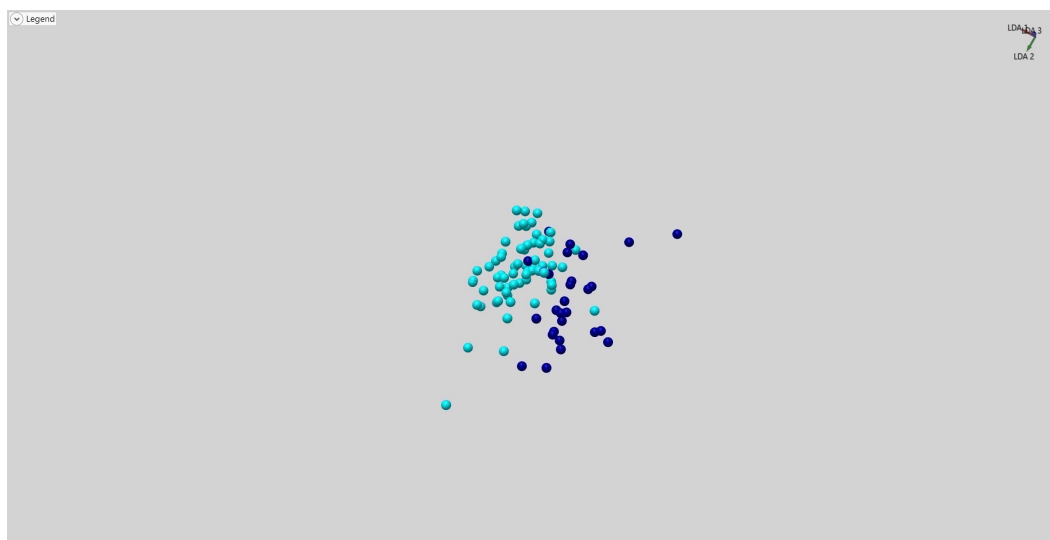


Figure 149. PCA plot of low grade and high grade REIMS spectra. IPA, m/z 600-800

Low grade (dark blue) and high grade (light blue) tumours segregate on LDA showing that the mass spectra are different.

This is the crudest form of tumour grade separation although it is the most important clinically, as high grade tumours require further treatment after surgery in the form of chemotherapy and radiotherapy and low grade tumours do not.

I performed supervised analysis using LDA. This showed good clustering of samples between the 4 WHO grades as can be seen in Figure 150.

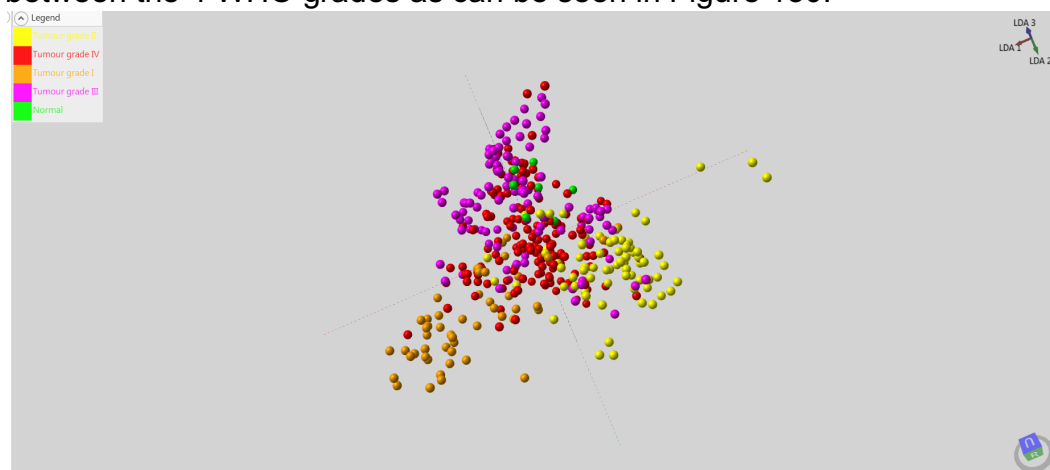


Figure 150. LDA plot of WHO grades from non-IPA and IPA REIMS spectra.

Good separation of the 4 different WHO grades is seen. WHO grade 1 (orange), grade 2 (yellow), grade 3 (pink) and grade 4 (red) with normal tissue in green.

Analysing data from the use of IPA solvent only produced improved class separation. For example, using equally matched numbers of samples from randomly chosen Grade 3 and Grade 4 spectra excellent separation is observed in Figure 151.

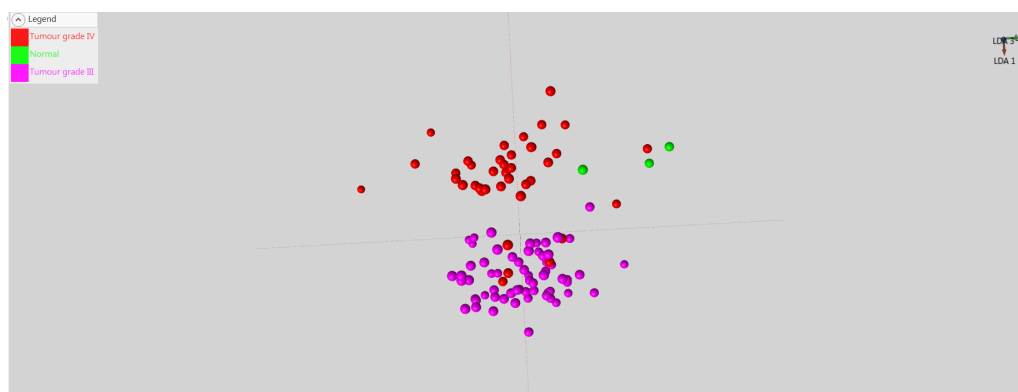


Figure 151. LDA plot of WHO grades from IPA REIMS spectra.

Good separation of the WHO grades 3 and 4 is seen. WHO grade 3 (pink) and grade 4 (red) with normal tissue in green.

PCA models were then constructed with the data as part of an unsupervised approach, with the knowledge that IPA only data may exhibit better separation. The data was first analysed with all spectra (IPA and non-IPA) as part of the dataset as shown in Figure 152.

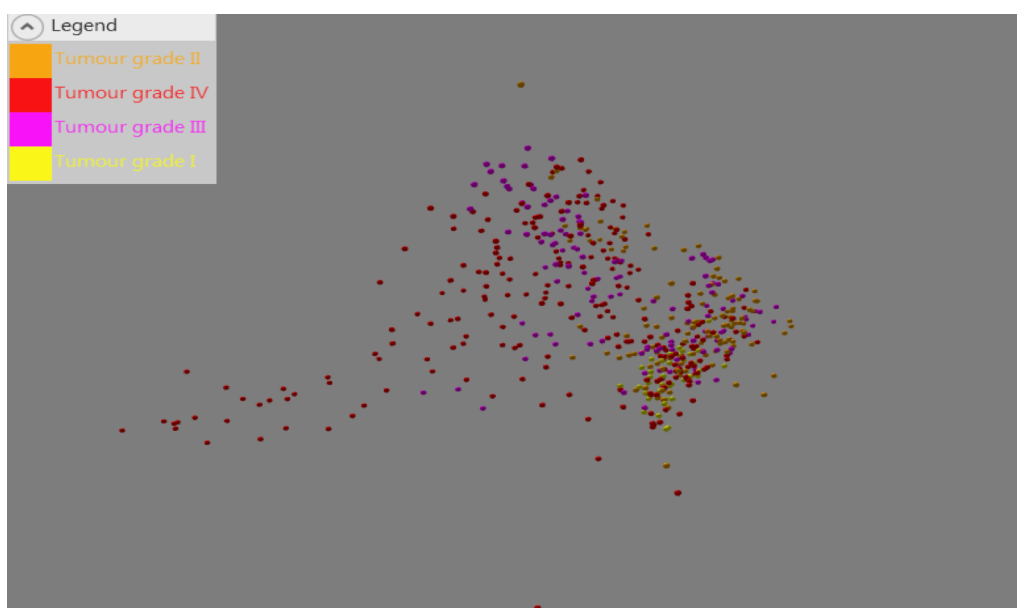


Figure 152. PCA plot using 300 components of tumour grade, REIMS IPA and non-IPA combined.

There is some separation of WHO grades: grade 1 (yellow), grade 2 (orange) grade 3 (pink) and grade 4 (red).

LDA plot of the same data shows good clustering as illustrated in Figure 153.

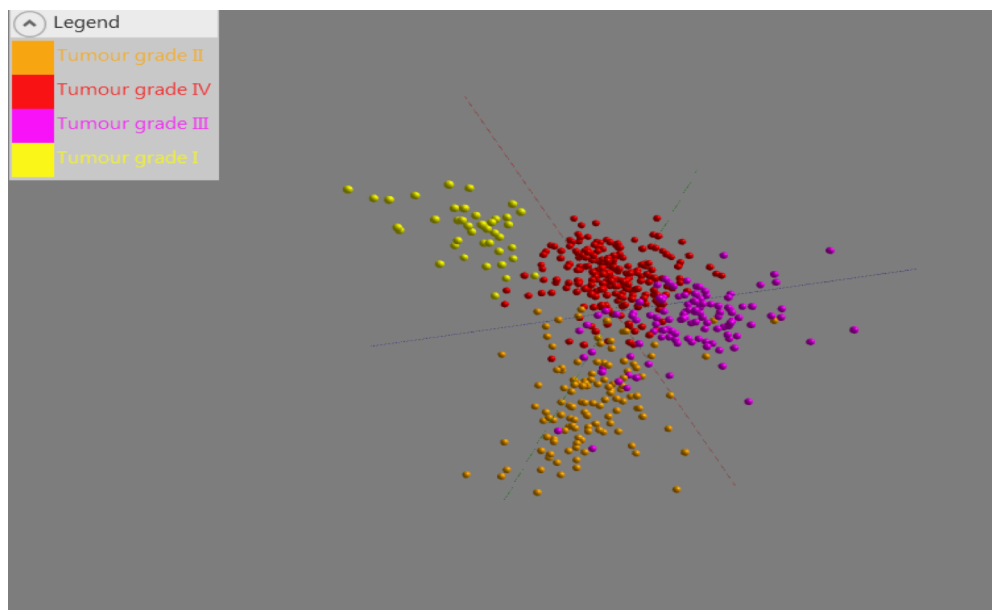


Figure 153. LDA plot using 300 components of tumour grade, REIMS IPA and non-IPA combined.

LDA shows better clustering of WHO grades: grade 1 (yellow), grade 2 (orange) grade 3 (pink) and grade 4 (red).

This data shows excellent separation on LDA and reasonable separation on PCA. This suggests a PCA-LDA model could perform well. When I used Recognition software to judge model performance on tissue spectra from patients not used to build the model the following results were obtained:

- 39 patient data files out of a total 46 patients were run in recognition
- 16 files classified 100% accurately
- A further 4 patients classified over 80% of the spectra correctly
- The most misclassifications were due to incorrect classification of Grade III as Grade II or IV (or vice versa)
- This recognition model correctly classified 100% of the spectra that were collected using IPA (n=16)

Hence IPA only data was analysed in the complete dataset of 75 patients. Comparing Mets, Radionecrosis and Normal very tight segregation is obtained on LDA as shown in Figure 154.

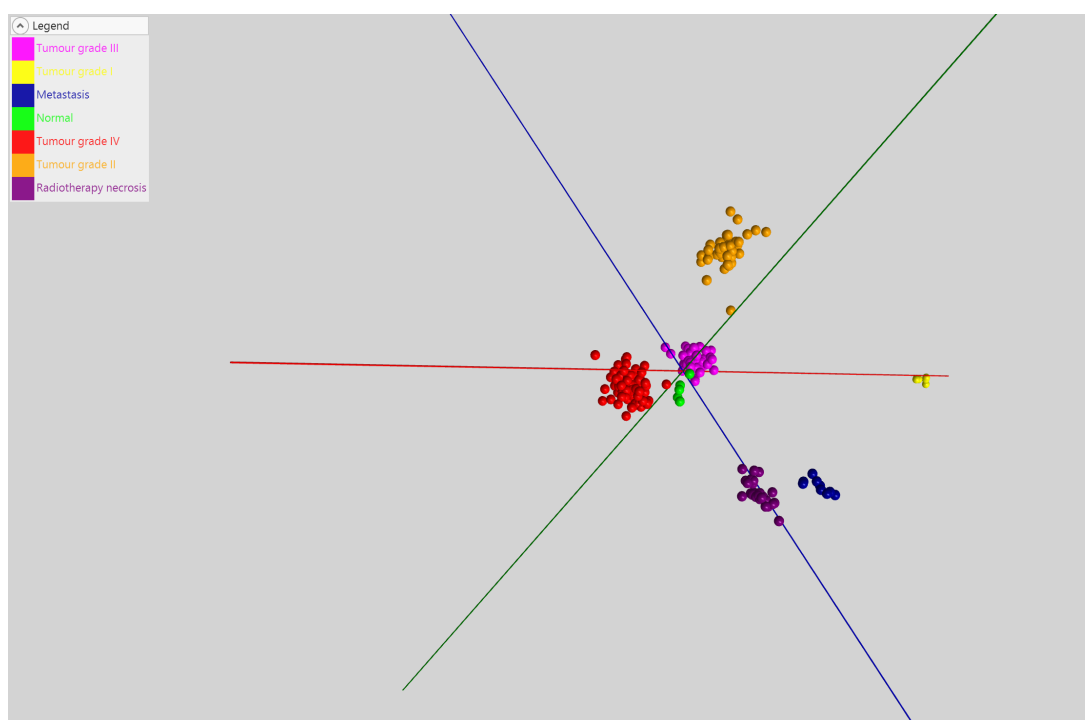


Figure 154. LDA plot of tumour grade for IPA results only, m/z range 100-1000. Excellent complete separation of WHO grades is seen, with clear separation of spectra obtained from radionecrotic tissue and metastasis.

Analysis of the PCA-LDA model loadings suggest that only the higher mass range 700-900 is required to produce this degree of discrimination between classes, shown in Figure 155.

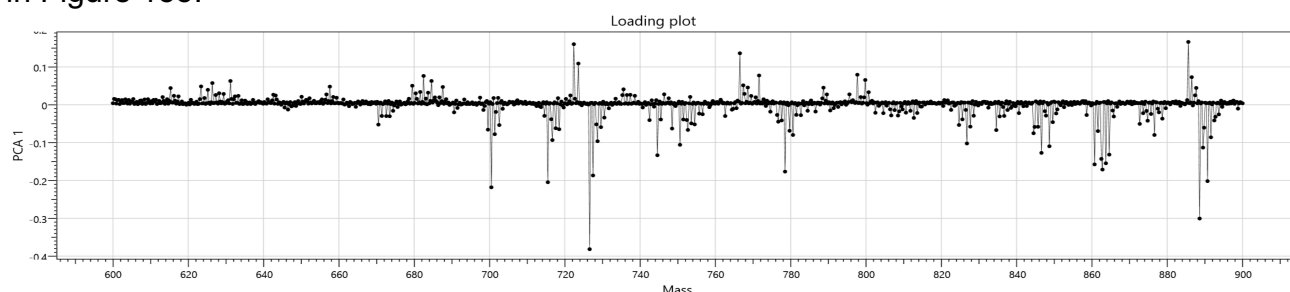


Figure 155. Tumour grade PCA-LDA loadings for the IPA model. The loadings from the PCA analysis show that the significant mass ranges which contribute most to the differences between WHO grades are from 700-900.

The cross validation result for low vs high grade model is shown in Table 32. This shows a low accuracy of 61% using LOPOCV.

Group result:

Group	Number of spectra	Number of passes	Number of failures	Number of outliers	Correct Classification Rate	
					Excluding outliers	Including outliers
10	4	4	0	0	100.00%	100.00%
28	27	7	20	0	25.93%	25.93%
35	7	5	2	0	71.43%	71.43%
36	18	1	17	0	5.56%	5.56%
37	2	0	2	0	0.00%	0.00%
38	21	8	13	0	38.10%	38.10%
39	52	43	9	0	82.69%	82.69%
40	5	5	0	0	100.00%	100.00%
42	29	14	15	0	48.28%	48.28%
43	36	34	2	0	94.44%	94.44%
44	1	1	0	0	100.00%	100.00%
45	30	28	2	0	93.33%	93.33%
46	7	7	0	0	100.00%	100.00%
47	6	6	0	0	100.00%	100.00%
48	15	15	0	0	100.00%	100.00%
49	4	4	0	0	100.00%	100.00%
50	9	9	0	0	100.00%	100.00%
53	23	0	23	0	0.00%	0.00%
54	10	7	3	0	70.00%	70.00%
55	9	0	9	0	0.00%	0.00%
56	3	0	3	0	0.00%	0.00%
57	8	2	6	0	25.00%	25.00%
58	11	11	0	0	100.00%	100.00%
60	5	2	3	0	40.00%	40.00%
61	2	1	1	0	50.00%	50.00%
63	5	4	1	0	80.00%	80.00%
64	24	13	11	0	54.17%	54.17%
65	1	1	0	0	100.00%	100.00%
66	9	0	9	0	0.00%	0.00%
67	8	8	0	0	100.00%	100.00%
68	5	1	4	0	20.00%	20.00%
69	6	0	6	0	0.00%	0.00%
7	9	9	0	0	100.00%	100.00%
70	21	0	21	0	0.00%	0.00%
71	21	19	2	0	90.48%	90.48%
72	8	8	0	0	100.00%	100.00%
73	19	17	1	1	94.44%	89.47%
Total	480	294	185	1	61.38%	61.25%

Confusion matrix:

	ii	iv	Outlier	Total
ii	74	48	1	123
iv	137	220	0	357
Total	211	268	1	480

Table 32. LOPOCV of tumour grade low vs high PCA-LDA model, non-IPA, m/z 600-800.

The cross validation results for all 4 WHO grades are shown in Table 33. The accuracy of the model was: 79.8% when excluding grade 3 from the data and 65% when including grade 3 in the dataset.

Group result:

Group	Number of spectra	Number of passes	Number of failures	Number of outliers	Correct Classification Rate
8	20	18	2	0	90%
29	5	1	4	0	20%
33	12	9	3	0	75%
1	1	1	0	0	100%
3	1	0	1	0	0%
10	8	8	0	0	100%
14	6	5	1	0	83.33334%
18	6	2	4	0	33.33333%
22	2	1	1	0	50%
21	4	4	0	0	100%
7	13	5	8	0	38.46154%
20	4	1	3	0	25%
24	8	3	5	0	37.5%
4	5	5	0	0	100%
12	6	1	5	0	16.66667%
19	6	4	2	0	66.66666%
31	6	6	0	0	100%
34	5	3	2	0	60%
Total	118	77	41	0	65.25423%

Cross-validation matrix:

	Tumour grade I	Tumour grade II	Tumour grade III	Tumour grade IV	Total
Tumour grade I	28	0	3	6	37
Tumour grade II	0	21	6	1	28
Tumour grade III	2	6	9	8	25
Tumour grade IV	1	4	4	19	28

Table 33. LOPOCV of tumour grade PCA-LDA model, non-IPA, m/z 600-800.
Grade 3 proves the hardest to correctly classify using this model. All other grades classify well.

5.5.8.3. Conclusion

The low grade vs high grade model gives an accuracy of 61.38%. Classification using all 4 WHO grades is possible. When excluding grade 3 accuracy for the 4 grade model was 79.8%, including grade 3 spectra reduces it to 65%.

On closer inspection of the PCA/LDA charts it seems Grade 3 is divided into 2 groups- those similar to grade 2 and those similar to grade 4 hence causing the lower cross validation results when included in the analysis. Interestingly, this reflects the clinical behaviour of grade 3 tumours, with some adopting a low grade course and others a high grade clinical course. More longitudinal clinical outcome data needs to be collected to verify if this is the case.

Using IPA data only grade 3 and 4 exhibited better separation: this is of high clinical utility in determining diagnostic biopsies and tissue grading in low grade to high grade transforming disease.

5.5.9. Genomic Studies

5.5.9.1. Introduction

The main questions to answer from data experiments were:

1. Is there a difference in REIMS spectra from mutated vs non mutated tissue of different key driver mutations in glioma?
2. Can models be built to learn the difference from mutated vs no non mutated tissue?
3. Can these models predict the mutational status of new samples of tissue with accuracy?

The hypothesis was that REIMS data should vary depending on which genetic mutations are causing the tumour phenotype as defined by immunohistochemistry. Founder mutations have downstream effects on tumour cell metabolism and lipidomics which should produce phospholipid changes detectable using the iKnife in surgery.

All histopathological samples were processed for an array of genetic mutations found in brain tumours as part of routine clinical practice. The genetic abnormalities tested include IDH-1 mutation, loss of ATRX expression and MGMT methylation. IPA and non-IPA data was analysed for these mutations. A limiting factor was subgrouping of the data, hence a range of cross over validation techniques were used. For some mutations (for example, non-IPA ATRX loss of expression) the numbers were so small that LOPOCV resulted in erroneous results.

5.5.9.2. IDH

234 spectra were collected from 118 tissue samples from 18 patients. LDA analysis of IDH mutation in GBM patients produces excellent separation of data as shown in Figure 156.

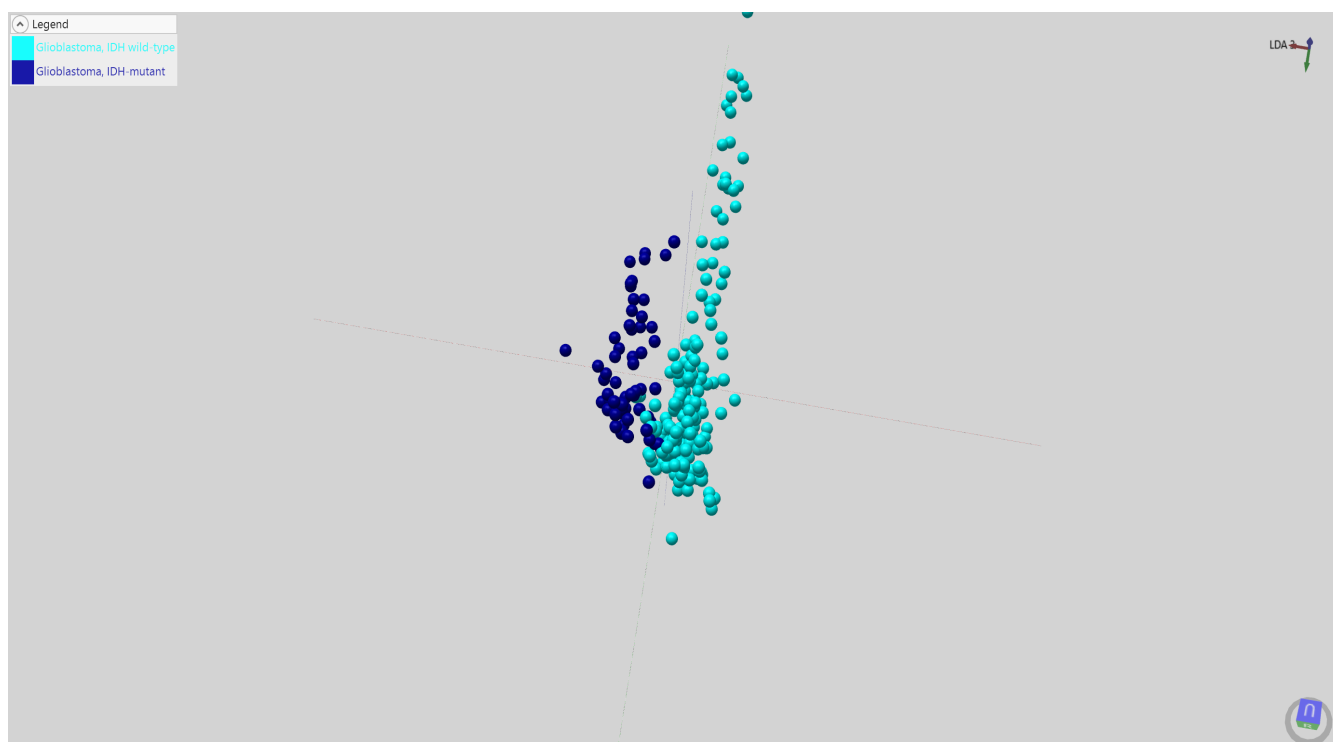


Figure 156. LDA plot of IDH mutation in GBM, non-IPA, m/z100-1000.
Excellent separation is seen between IDH mutated (dark blue) and non-mutated (light blue) spectra in GBM using non-IPA data.

Cross validation was good reaching 78.6% accuracy (75.6% sensitivity and 82.9% specificity) and as shown in Table 33.

Confusion matrix:

	0	1	Outlier	Total
0	39	5	3	47
1	37	115	35	187
Total	76	120	38	234

Table 33. Leave 20% out cross validation of the IDH mutation REIMS model.
Classification was achieved using non-IPA data reaching 95.8% sensitivity.

Good separation between IDH mutant and non-mutant spectra from all tumour grades is also seen when using IPA, as shown in Figure 157.

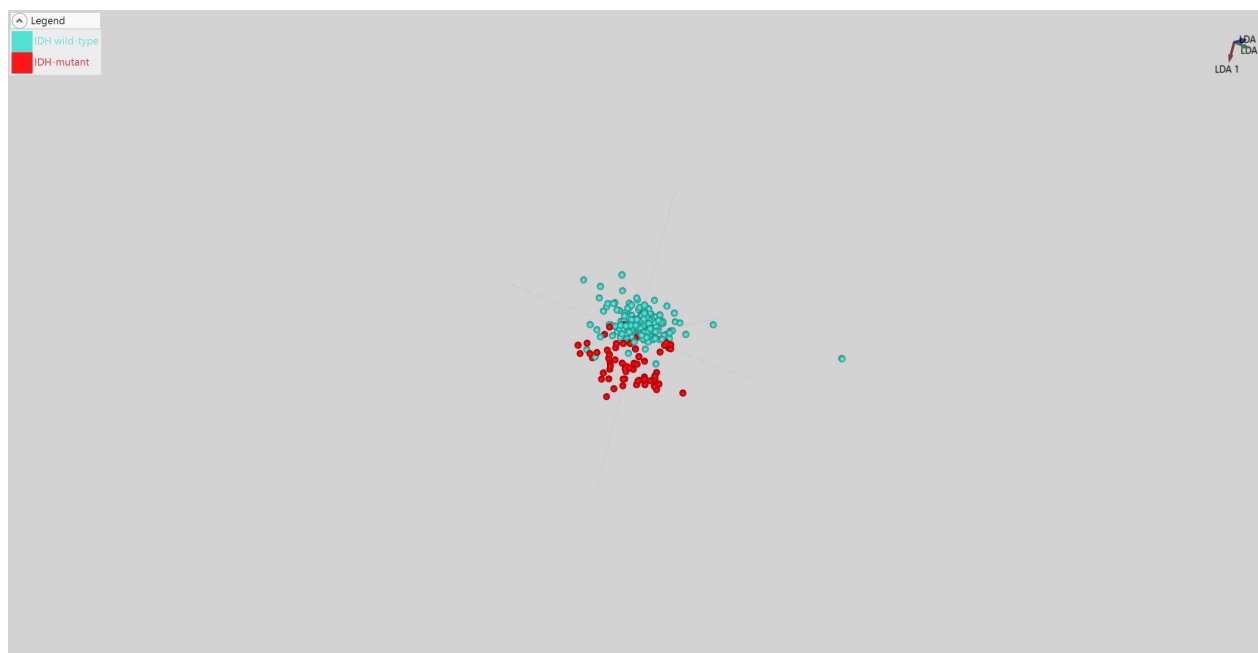


Figure 157. LDA plot of IPA data and IDH mutation status using REIMS.
Excellent separation is seen of IDH mutant (red) and non-mutant (green) spectra from all tumour samples.

The cross validation for the IPA model showed this to be weaker than non-IPA: with an accuracy of 59.4% when using leave 20% out cross validation (Table 35).

Confusion matrix:

	0	1	Outlier	Total
0	97	86	33	216
1	13	48	9	70
Total	110	134	42	286

Table 35. Leave 20% out cross validation on the PCA-LDA model for IDH mutation using IPA data.

When IPA and non-IPA data is combined the PCA-LDA model performance is intermediate as shown in Table 36 reaching an accuracy of 71.1% on LOPOCV.

Group result:

Group	Number of spectra	Number of passes	Number of failures	Number of outliers	Correct Classification Rate	
					Excluding outliers	Including outliers
10	27	26	1	0	96.30%	96.30%
13	1	1	0	0	100.00%	100.00%
14	38	38	0	0	100.00%	100.00%
16	27	17	10	0	62.96%	62.96%
18	7	7	0	0	100.00%	100.00%
19	8	0	8	0	0.00%	0.00%
2	18	17	1	0	94.44%	94.44%
21	11	11	0	0	100.00%	100.00%
22	1	1	0	0	100.00%	100.00%
23	15	15	0	0	100.00%	100.00%
24	20	19	1	0	95.00%	95.00%
28	30	7	23	0	23.33%	23.33%
3	5	0	5	0	0.00%	0.00%
35	7	7	0	0	100.00%	100.00%
36	18	18	0	0	100.00%	100.00%
37	2	1	1	0	50.00%	50.00%
38	21	1	20	0	4.76%	4.76%
39	52	48	4	0	92.31%	92.31%
4	15	0	15	0	0.00%	0.00%
42	30	9	21	0	30.00%	30.00%
43	36	34	2	0	94.44%	94.44%
44	1	1	0	0	100.00%	100.00%
45	30	30	0	0	100.00%	100.00%
46	7	3	4	0	42.86%	42.86%
47	6	3	3	0	50.00%	50.00%
48	15	15	0	0	100.00%	100.00%
49	4	4	0	0	100.00%	100.00%
5	16	0	16	0	0.00%	0.00%
50	9	8	1	0	88.89%	88.89%
53	3	3	0	0	100.00%	100.00%
6	1	1	0	0	100.00%	100.00%
65	1	0	1	0	0.00%	0.00%
68	5	2	3	0	40.00%	40.00%
7	33	23	10	0	69.70%	69.70%
Total	520	370	150	0	71.15%	71.15%

Confusion matrix:

	0	1	Total
0	185	78	263
1	72	185	257
Total	257	263	520

Table 36. Leave one patient out cross validation results of the combined IPA and non-IPA REIMS model for IDH mutation.

The loadings plot in Figure 158 shows that the lower m/z ranges are most important for discrimination between IDH WT and IDH MT spectra. Individual peak analysis of the first principle component shows that m/z 255 is the most significant separator, followed by 283 and 281.

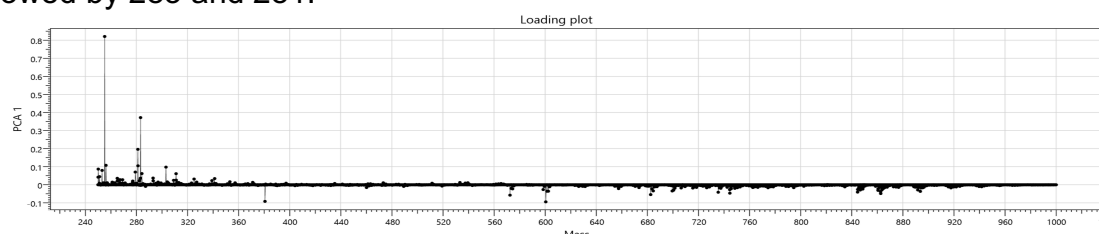


Figure 158. The loadings plot for the IDH mutation model using REIMS data.

Improved classification performance was seen in distinguishing IDH mutated from non-mutated spectra in GBM.

259 spectra were analysed consisting of 194 GBM WT and 65 GBM MT from 140

tissue samples from 18 patients. IPA and non-IPA data was combined with a m/z range of 500-1000.

The cross validation accuracy of the PCA-LDA model using 20 percent out reached an impressive 94% (Table 37). LOPOCV only reached 50%.

Confusion matrix:

	glioblastoma, idh wild-type	glioblastoma, idh-mutant	Outlier	Total
glioblastoma, idh wild-type	185	7	2	194
glioblastoma, idh-mutant	8	56	1	65
Total	193	63	3	259

LOPOCV is poor at 50%

Confusion matrix:

	glioblastoma, idh wild-type	glioblastoma, idh-mutant	Outlier	Total
glioblastoma, idh wild-type	129	64	1	194
glioblastoma, idh-mutant	63	2	0	65
Total	192	66	1	259

Table 37. Leave 20% out cross validation (top) and leave one patient out cross validation (bottom) for IDH mutation in GBM patients using REIMS data.

5.5.9.3. ATRX

454 spectra were collected (290 ATRX retained, 164 ATRX loss) from 224 tissue samples and 47 patients. IPA and non-IPA datasets were combined, lockmass of 544 was used with one burn per spectrum with background subtraction, normalisation and a mass range of 600-1000.

The PCA-LDA model generated was cross validated using LOPOCV and leave 20% out cross validation. LOPOCV reveals an accuracy of 70% a sensitivity of 90% but a specificity of only 50%. The model correctly classifies nearly all ATRX mutated spectra but misclassifies 50% of the ATRX normal spectra as ATRX mutated (Table 38).

Confusion matrix:

	0	1	Outlier	Total
0	109	109	72	290
1	1	148	15	164
Total	110	257	87	454

Table 38. ATRX cross validation using the PCA-LDA model for REIMS tumour data.

In order to balance for this, 120 ATRX retained spectra were randomly removed to better balance the training dataset as it seemed to be over classifying for the over represented group. This matched ATRX model combined IPA and non-IPA data and consisted of 342 spectra from 162 tissue samples from 32 patients.

LOPOCV and leave 20% cross validation improved as shown in Table 39 with

reduced misclassification of non-mutated spectra. LOPOCV gives an accuracy of 64.6% and leave 20% out gives an accuracy of 75% a sensitivity of 84.1% and a specificity of 66.9%.

LOPOCV

Confusion matrix:

	0	1	Total
0	104	74	178
1	47	117	164
Total	151	191	342

20 percent CV

Confusion matrix:

	0	1	Total
0	119	59	178
1	26	138	164
Total	145	197	342

Table 39. Cross validation results of ATRX mutation from REIMS data using a matched training set.

LOPOCV (top) and leave 20% cross validation (bottom) was performed on the PCA-LDA model.

5.5.9.4. MGMT

The best classifier for MGMT methylation used IPA data in the 600-1000 mass range. 75 tissue samples were included yielding 157 spectra consisting of 29 non-methylated, 128 methylated spectra from 15 patients.

The PCA-LDA model generated from the complete training set in OMB was cross validated using LOPOCV and reached an accuracy of 73.9%, a sensitivity of 81.2% and a specificity of 41.3% as shown in Table 40.

Confusion matrix:

	0	1	Total
0	12	17	29
1	24	104	128
Total	36	121	157

Table 40. Results of LOPOCV for the MGMT methylation REIMS PCA-LDA model.

Over classification of MGMT non-mutated tissue as MGMT mutated is seen in this model as the training dataset is biased for MGMT mutation (128 spectra vs 29). To mitigate this a more balanced training set was created to build the model on. Table 41 shows the results of cross validation on this model with improved accuracy and reduced misclassification. This resulted in a reduced accuracy of 60.2%, a sensitivity of 63% and an improved specificity of 55.1%.

Confusion matrix:

	0	1	Outlier	Total
0	16	7	6	29
1	9	34	11	54
Total	25	41	17	83

Table 41. LOPOCV for MGMT methylation using the modified REIMS PCA-LDA model.

5.5.9.5. Summary

I have shown that IDH shows strong clustering on LDA with some models performing well on cross validation. Sensitivity reached 75.6% with good specificity at 82.9%. Further data should improve the systems ability to detect IDH mutation and confirm whether the accuracy is artefactual or real. In particular my sensitivity of 96.9% and specificity of 96.3% of detecting IDH mutation in GBM suggests mass spectroscopy may be able to ascertain if GBM is primary or secondary.

ATRX mutation models did not perform well with a high degree of misclassification of normal spectra as mutated. This was reduced by producing a more matched training dataset but accuracy remained low at 75%, sensitivity at 84.1% but a disappointing specificity of 66.9%.

MGMT posed similar challenges in terms of accuracy. The best model produced an accuracy of 73.9%, a sensitivity of 81.2% and a specificity of 41.3%. Using a balanced dataset for training resulted in an improved specificity to 55.1% but reduced overall accuracy to 60.2%.

5.5.10. Conclusion

The data I have presented in this chapter shows that when REIMS is used during surgery as the iKnife recognisable mass spectra are observed in a reproducible manner. These spectra can be used to show differences between tissue types to answer key clinical questions. These differences can be detected using supervised and unsupervised analysis. PCA-LDA machine learning models can be successfully trained on the data obtained.

These models can be successfully used to predict the clinical characteristics of data new used for cross validation to find out if tissue samples are normal or tumour, the type of tumour, tumour grade, tumour density and the genetic mutational status of tissue being analysed.

6. Discussion

6.1. Introduction

This thesis is about successful translational research that has taken place in the NHS within an astonishingly small time frame of 3 years. Within these 3 years a project was started from scratch to bring innovative technology into first in human use and successful trials were designed, registered, opened and closed with recruitment to target. Novel data was analysed from *in vivo* sources using techniques which had never been used together before on the living human brain.

This work has brought to bear the immense knowledge base and traditions of analytical chemistry, data science, cancer biology and neurosurgery to help develop new avenues to diagnose and treat some of the deadliest and complex cancers that can afflict us. These disciplines of science have been combined to provide a solution to the problems faced daily by patients and their oncology teams. Central to this is the role of the surgeon who is at the start of the patient pathway and is the source of molecular data which drives all treatment decisions and will ultimately drive a revolution in cancer care.

The first section of this thesis outlined Raman spectroscopy and REIMS as molecular chemometric methods with potential in the operating theatre. The following sections outlined my work in developing the technologies for use in humans through to first use in surgery. The results sections have detailed the molecular data obtained from the 2 NIHR registered clinical trials which resulted from this thesis and have recruited to completion.

The following chapters will deal with the interpretation of the results with reference to the aims of the thesis. The limitations of the work will be discussed as well as future directions for further trials and development of the chemometric, data science and surgical methodology used in this thesis.

6.2. Contribution to the field of neurosurgical oncology

This chapter outlines the main contributions made to Neurosurgical oncology by this thesis. The main contribution has been to develop a novel methodology to help solve complex problems in glioma surgery. The particular combination of techniques in this thesis has not been applied to glioma surgery before.

All of the thesis aims are novel contributions to neurosurgery, namely:

1. To develop 2 new chemometric technologies for 'First in Human' use
2. To introduce a new paradigm for the *in vivo* modelling of brain cancer to advance research into this condition
3. To develop new methodologies for the real-time molecular profiling of brain tumours during surgery
4. To use the new technology, models and methodologies developed to answer key clinical questions: a platform for precision medicine

This proof of concept study was an IDEAL framework (McCulloch et al., 2009) stage 2a (development of an idea looking at technical and procedural success) and early stage 2b (exploring clinical indications of new technology) study. The work in this thesis may be an important contribution to national and international neurosurgical innovation and is a potential glimpse into the future of neurosurgery. It is a future where data heavy processing of molecular data in the operating theatre occurs instantly to both guide the surgeon and characterise the disease being operated on in more detail than ever before to identify next steps from a range of emerging treatment options.

6.2.1. Development of 2 new modalities for surgery

Raman spectroscopy and the iKnife were successfully used during surgery for a combined total of 75 patients. The fact that both clinical trials recruited to completion showed that it is possible to translate technology from the laboratory to the operating theatre within a timeframe of 3 years.

Both systems were supported by technical teams to trouble shoot first use in humans. Sterilisation guidelines were developed and closely followed for sterilising the Raman probe whereas for the iKnife all components used were sterile, disposable pieces of equipment.

Both systems were integrated into the workflow of the operating theatre and collected a vast amount of data totalling over 50GB. Towards the end of the study only a 20 minute increase in operating time was caused by using these technologies per case. Although there was some data loss as a result of annotation errors in samples this did not exceed 5% of the samples collected, with the data loss rate falling significantly to zero once the trial was established after 20 cases. System failure resulted in an inability to collect data on a total of 6 patients between both studies.

6.2.1.1. Raman

The main challenge for developing *in vivo* Raman spectroscopy was the exclusion of infrared light sources. As detailed in section 3.2.2.4 these sources were carefully measured when the operating theatre was in routine use and then systematically excluded using a checklist at the time of taking a Raman reading. This enabled Raman spectra to be measured successfully.

The safe use of the probe was dependent on correct sterilisation protocols being developed and followed, laser safety protocols being used at the time of use of the probe (eye protection, laser window blinds and laser notification outside the operating theatre) and microscope visualisation of the probe tip at all times during Raman spectrum acquisition. This last step required the operating microscope to be optically filtered for the wavelength of laser being used.

The system was successfully used for over 31 patients. In 1 patient data was not collected due to fibre damage to the probe from handling during sterilisation and in 2 patients an SSD memory card connected to the CCD failed requiring replacement. In 2 cases the system froze requiring rebooting without loss of any spectroscopic data. Such events are to be anticipated when using new technology and at no point were patients put at risk.

The surgeon experience whilst using Raman spectroscopy was also very important. 3 surgeons used the system. They found it to be a tool which was intuitive to use as a 'point and shoot' system. There was a learning curve with using the system, mainly to do with the distance away from tissue the probe needed to be held and with the fact that for all surgeons this was the first time they had used a laser during surgery. The system was found to be rapid and easy to use with the footswitch trigger. Indeed, it was so fast and easy to use that it was used to interrogate different areas of the resection cavity repeatedly. The fact that the analysis is non-destructive lowered the threshold for use even further as there wasn't the fear of potentially destroying normal tissue when seeking reassurance at the end of surgery. This method of use seems to have been the most common: that is, use of the tool at the end of resection using current technology to identify areas of possible residual tumour. The decision would then be up to the surgeon whether to resect further tissue if deemed possible.

By the end of the study the Raman cart was a regular fixture in the operating theatre, occupying a place next to our neuronavigation system and operating microscope. This exemplifies its journey from the chemistry laboratory right into patient care.

6.2.1.2. iKnife

The big obstacle for introducing the iKnife into neurosurgical use was the Signal to Noise Ratio (SNR). Some data was discarded because of non-diagnostic data being acquired in the triglyceride range rather than higher mass ranges. In some cases, this was due to the incorrect bipolar electrical generator settings, in other cases it was due to incorrect tissue burning using the bipolar forceps. A strict regime of checks prior to each burn and after each burn reduced this risk, together with repeated unblocking of the suction tubing connecting the bipolar forceps with the atmospheric interface of the iKnife. Although time consuming this did lead to data in over 75 patients being collected successfully and the system became a regular site in the operating theatres.

This system therefore had a steeper learning curve than Raman. Due to the SNR

difficulties, IPA was developed from patient 36 onwards as a solution to improve signal. This did however shift some peaks making direct comparison of IPA and non-IPA data difficult. After this the system had improved stability and data loss was reduced. Tube blockage was another factor which was noticed in the first 10 cases. A tube blockage protocol was introduced whereby the forceps and tubing were checked, flushed and tubing sections cut as escalating steps to ensure signal could reach the iKnife machine as quickly as possible.

Equipment failure prevented data collection from 3 patients. In 1 case the system had been shut down inadvertently by theatre staff in the morning not leaving enough time for the vacuum to form before the surgical case started. The stepwave component of the machine failed resulting in the need for servicing whilst 2 cases were operated on. Theatre time prevented data collection from 2 cases. In these cases, surgery was complicated and required more time than usual precluding the use of the iKnife as there was no time to use it.

There were no equipment failures encountered during data collection. Once the system was setup and had passed setup checks it reliably produced data. There were no sterilisation issues as components were disposable.

The surgeon experience of the iKnife was positive. The system is built into bipolar forceps which are widely used in neurosurgery so using the system was very intuitive. Frequent forceps blockage did slow down data collection, to the point that 2 bipolars had to be prepared for each case (one for routine use, one for data collection). There also seemed to be a longer time delay in getting a signal compared to Raman but this was only by a few seconds. There were minimal safety issues as no lasers were used and all aspects of equipment were familiar to the surgeon and already used in surgery. Operation was simple via a foot pedal.

6.2.2. Concept and methodology for real time molecular profiling

The concept of real time molecular profiling established in this thesis is completely new for surgery in glioma. Although a range of techniques have been used on *ex vivo* tissue including Raman, mass spectrometry and various optical methods such as confocal microscopy as discussed in section 2.9, the use of *in vivo* techniques has been limited to one study which was published during the time this thesis was being conducted (Jermyn et al., 2015b).

The work in this thesis has contributed a novel platform and methodology for obtaining *in vivo* molecular data in a fast, reliable way with minimal workflow disruption. The overarching concept can be coined 'molecular guided surgery' as molecular information can be used during surgery to define tissue to help with surgical decision making. With regards to this concept particular original contributions have been made with respect to the data source, data collection, model generation and model testing.

6.2.2.1. Data source

A major contribution has been the development of 2 new sources of molecular information during surgery. Raman spectroscopy and mass spectrometry have been developed into brain specific systems capable of yielding rapid *in vivo* chemical information. These methods have been developed in this thesis to be safe, reliable, repeatable and surgically usable and so the development of Raman spectroscopy and REIMS specifically for neuro-oncological surgery serves in itself as a major contribution to the field as no previous examples exist of these technologies being used in neurosurgery in Europe.

6.2.2.2. Data collection

Matching the molecular information obtained using the 2 experimental technologies with key clinical histopathological features was a crucial step in allowing associations to be discovered. An *in vivo* model of integrating molecular information, 3D neuronavigation positional information and histopathological information was developed as an original contribution to the field (see section 3.5.3). The specific system of anatomical specimen annotation (see Appendix 2, section 8.2) enabled fast, accurate labelling of multiple specimens as documented in the data entry sheet in Appendix 1, section 8.1.

The aim was to combine spectroscopic data with clinical data from histopathology and positional data from neuronavigational MRI images, and eventually further data can be added to this model in terms of clinical outcome and treatment modalities. The platform is flexible and can be used to investigate any new technology which produces information from point analysis of tissue. Indeed, this system has already been modified and has enabled related *in vivo* studies on the human brain (Grech-Sollars et al., 2016).

6.2.2.3. Model generation

The new data collected in this thesis is very different from previous molecular data paradigms as it is related to modern analytical chemistry, not molecular biology directly. The complex data sets consist of millions of fragments of molecules (mass spectrometry) or the information relating to chemical bonds superimposed from the many compounds that make up human tissue (Raman spectroscopy).

These complex data sets have required non-standard methods of analysis in order to find the complex non-linear relationships between chemical data and clinical information. Developing multivariate models on this data using machine learning has been a central task of this thesis and the data methodology used serves as an original contribution to the processing of *in vivo* brain spectroscopic data. In particular models using PLS-DA, SVM, PCA-LDA were found to perform the best with *in vivo* data, with less reliance on Random Forrest and neural networks than other studies on *ex vivo* tissue (Eberlin et al., 2012) (Azimi et al., 2015) (Yang et al.,

2015).

The iterative and exploratory approach to developing balanced comparisons, discovering data effects and building robust models serves as important work for future machine learning approaches to tissue classification and outcome prediction in surgery. In particular the data obtained allows more accurate predictions of numbers of patients and numbers of tissue samples required to train models for tissue recognition. This is crucial as patients and brain tissue samples are expensive resources in clinical trials. Selecting tissue efficient models has obvious value in this context.

6.2.2.4. Model testing

The methodology for testing the accuracy of models with cross validation is an important contribution to the field. A range of techniques were used such as leave one out cross validation, leave 20% out cross validation and leave one patient out cross validation. The majority of studies published in the field use leave one out cross validation which was found in this thesis to be susceptible to over-fitting. The accuracies obtained from LOOCV were not replicated when using the other methods of cross validation indicating that by building a model on all but one spectrum individual patient batch effects dominated the decision making in the model. A recommendation has been the use of leave 20% out or leave one patient out cross validation for models to perform well *in vivo*.

When selected appropriately models are able to predict tissue characteristics in real time with acceptable accuracy. The 3D reconstruction in real time of margin information in section 3.7.9 shows this can be done fast enough to inform the surgeon and serves as a major contribution to the field.

6.2.3. A new platform for precision medicine

The work in this thesis has developed a new platform for molecular level decision making. The data derived can be used to create models to detect tumour tissue, tumour grade, density and tumour type. However, a major contribution is the ability to detect tissue containing key drivers of gliomagenesis which are used as predictive and prognostic biomarkers. This allows the technology to inform further treatment options for the patient and so allows the platform to be one of the first examples of precision medicine in neurosurgery.

Real time *in vivo* detection of key molecular phenotypes in humans offers the opportunity to gain a unique insight into tumour biology. The ability to spatially map this information with neuronavigation as described here makes Raman spectroscopy a potential tool for investigating *in vivo* clonal evolution and spatiotemporal genomic architecture within the living patient. This could represent the most accurate model of this complex disease for the understanding of key drivers of tumour pathogenesis *in situ*. We have identified tissue containing early driver mutations in glioma, but it is

likely that Raman spectroscopy will be able to obtain a much more extensive molecular profile of disease in larger multicentre trials. Through providing molecular information in a clinically useful timeframe the rapid molecular analysis of tumours using Raman spectroscopy may be pivotal to the development of a precision medicine approach to cancer.

6.3. *in vivo* spectroscopy for the Human brain

The molecular biology of tumours underlies the development of novel treatments and is becoming an integral part of clinical decision making. Medulloblastoma, the commonest brain tumour in childhood, until recently was treated as a single entity but actually consists of 4 distinct molecular subtypes (Northcott et al., 2011) with different outcomes and emerging targeted treatments (Kieran, 2014). Current molecular classification systems are not complete and more effective markers for diagnosis, prognosis and guiding precision therapy remain to be discovered. The early diagnosis and molecular subtyping of brain tumours have therefore been suggested as national research priorities leading to better outcomes and prognosis (Department of Health and Social Care, 2018) whilst recognising that it is challenging getting appropriate test results within a clinically meaningful timeframe (Kurian et al.). Intraoperative spectroscopy offers a solution to these key problems facing neuro-oncology.

The work undertaken in this thesis shows that *in vivo* spectroscopy for the human brain is possible. The accuracy and quality of the spectroscopic data produced allow a range of conclusions to be drawn. In the following sections the results from Raman spectroscopy and the iKnife will be discussed with reference to the literature.

6.3.1. Raman spectroscopy

It is known that Raman spectra contain detailed systems-level information spanning genomic, proteomic and lipidomic domains. The successful analysis of this information to provide key elements of multiple types of clinically useful information for brain tumours has not been performed before. In this thesis I have demonstrated for the first time, to the best of my knowledge, that all of the following can be defined using Raman spectroscopy in the human brain:

- Presence of tumour
- Tumour density
- Tumour type
- Tumour grade
- Tumour margin
- Genetic status

Furthermore, it has been demonstrated in this thesis that *in vivo* data can be used to

construct these models. Figure 42 is reproduced below and shows that excellent quality *in vivo* Raman spectra were obtained by our Raman system. The only published study of *in vivo* Raman spectroscopy in brain tumours (Jermyn et al., 2015a) illustrates the difficulty in getting good quality spectra without interference from fluorescence (Figure 159). By simple visual comparison between data from both studies it can be seen that data from our Raman system contains clear peaks as opposed to the over averaged and smoothed data from Jermyn *et al.* It is uncertain how machine learning algorithms can gather enough data from the appearances of spectra like in Figure 159 and classify tissue accurately. Indeed the machine learning algorithms used by Jermyn *et al.* were not specified in their publication.

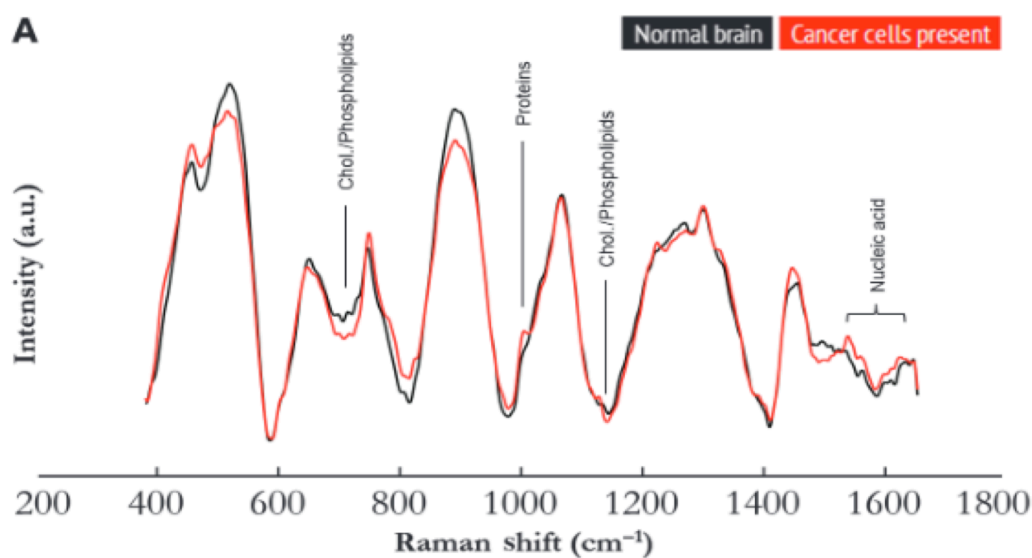


Figure 159. Averaged Raman spectra from the study by Jermyn et al. (Jermyn et al., 2015a)

This Raman spectrum shows very few peaks a lot of background ambient light. It is not clear how diagnostic information can be obtained from such a featureless spectrum.

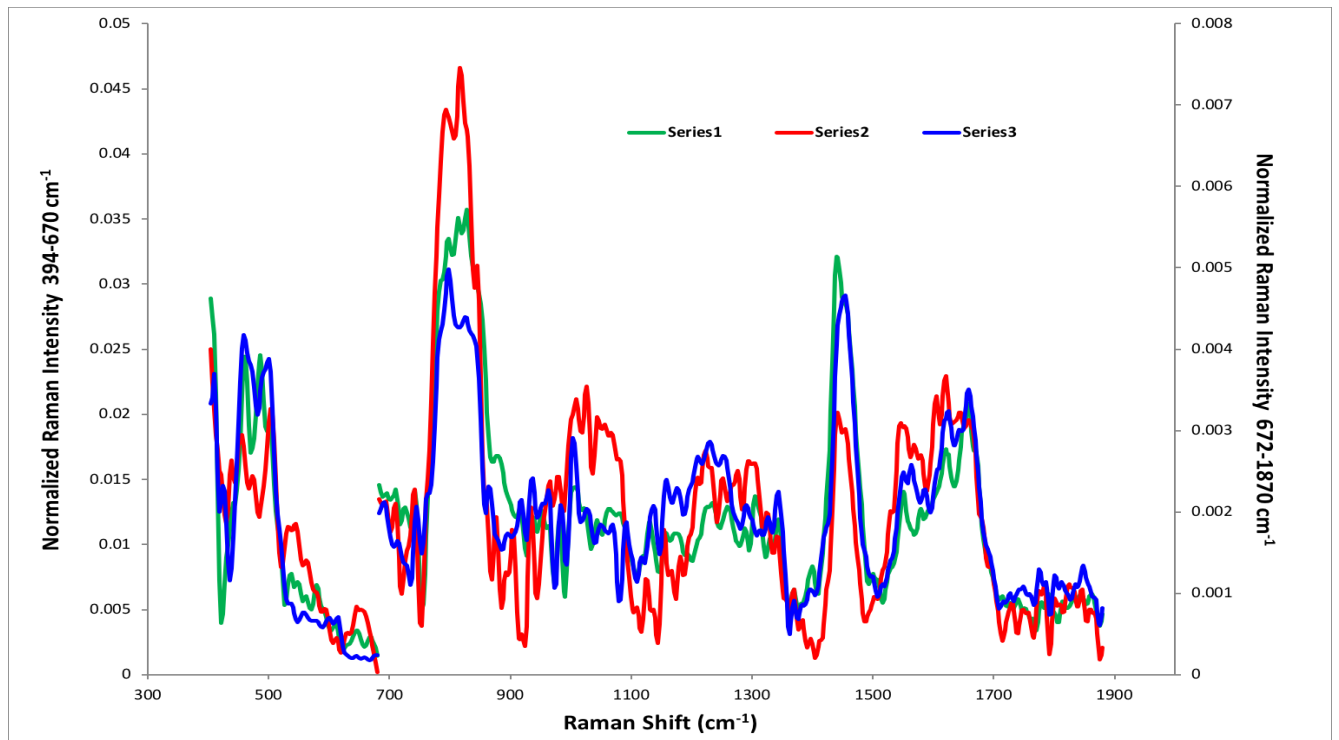


Figure 42. Raman spectrum of in vivo brain tissue of different WHO grades.

Example Raman spectra from range 394-1870 wave numbers. A break at 700nm is seen due to optical filters used in the system. Series 1=Grade 2, Series 2-Grade 3 and Series 3-Grade 4 tumours on the WHO grading scale.

Indeed, it seems that the *in vivo* spectra obtained in this thesis bear more similarities with published *ex vivo* studies on human brain tumour tissue, suggesting that our system has obtained true high quality Raman spectra free from the majority of background contaminating IR light found in the operating room. Figure 160 from Amharref *et al* shows a number of key peaks which are obvious and shared with the spectra obtained in this thesis. In particular the following peaks:

1658
1436
826

The peaks in the 1600 and 1400 range are also shared with other *ex vivo* studies, such as Leslie *et al.* (Leslie *et al.*, 2012) shown in Figure 161 below.

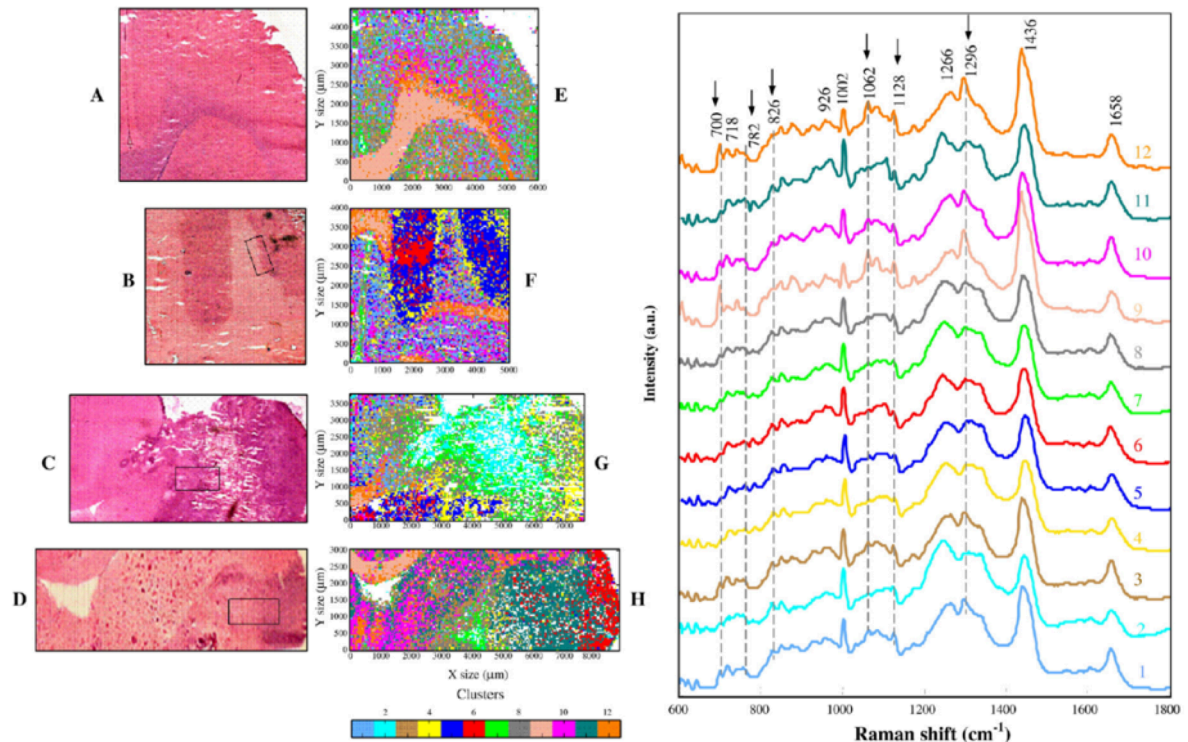


Figure 160. Photomicrographs (H&E staining) of healthy (A) and glioma (B–D) brain tissue sections.

Pseudocolour Raman maps E–H are based on 12-means cluster analysis on sections A–D, respectively. As per contra, representative cluster-averaged Raman spectra collected from healthy and glioma brain tissue sections. Spectra are shown with the same colour than in the pseudocolour maps E–H. Reproduced from Amharref et al, 2007 with permission from (Amharref et al., 2007)

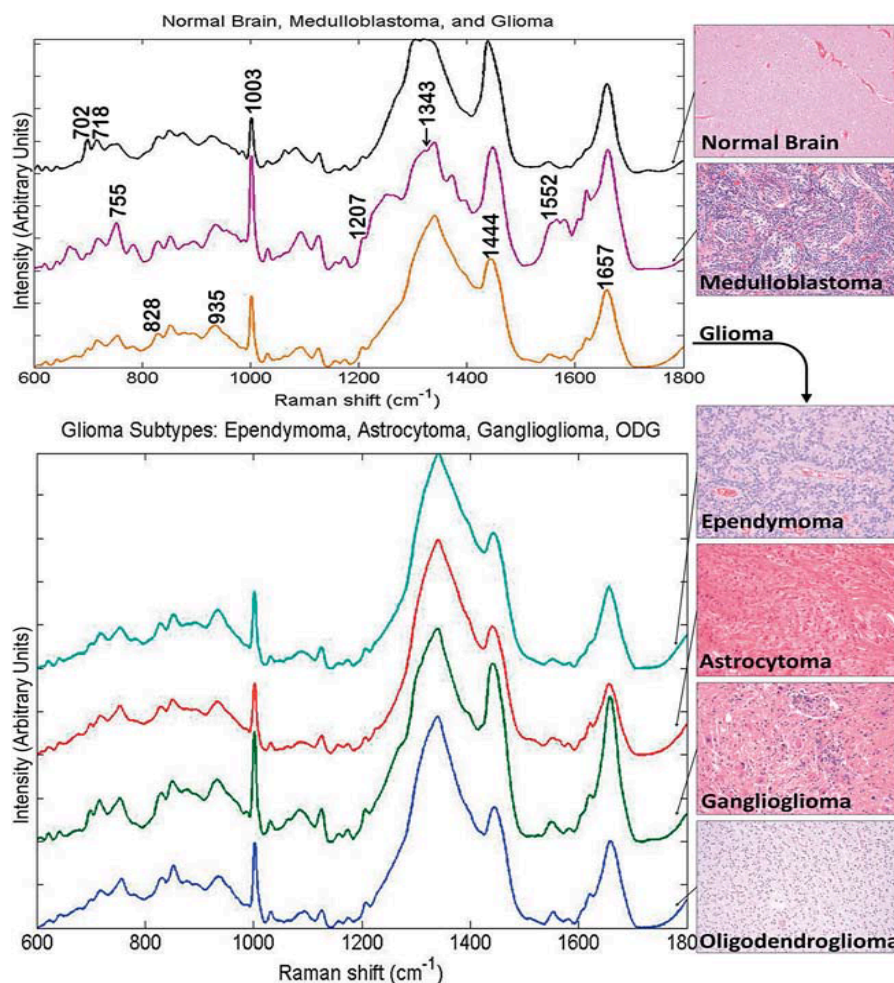


Figure 161. The mean *ex vivo* Raman spectra of normal tissue, medulloblastoma and glioma tissue (top) and mean spectra for each type of tumour (bottom).

Corresponding H&E slides of the tissue samples are shown on the right. Key peaks are labelled. From Leslie et al. (Leslie et al., 2012).

Based on other *ex vivo* work in the field, the significant peaks are summarised by Leslie et al. (Leslie et al., 2012) in Figure 162.

Peak	What it distinguishes	Metabolite(s)
702	Normal, astrocytoma and ganglioglioma	Cholesterol/cholesterol ester
718	Normal, astrocytoma and ganglioglioma	Lipids (sphingomyelin)
1,003	All, strongest in medulloblastoma	Phenylalanine, protein, hemoglobin
1,207	Medulloblastoma	Collagen, tyrosine
1,343	Medulloblastoma (valley)	Nucleic acid
1,552	Medulloblastoma	Hemoglobin, tryptophan
828	Glioma	Proline, hydroxyproline, phosphatidylserine
935	Glioma	Collagen
755	Glioma and medulloblastoma	Phosphatidylethanolamine, tryptophan, hemoglobin
1,444	Glioma and medulloblastoma	Phosphatidylserine, cholesterol, sphingomyelin
1,657	Glioma and medulloblastoma	Water, protein

Figure 162. Significant Raman spectral characteristics for brain tumours-reproduced from Leslie et al. (Leslie et al., 2012)

Using data from Figure 162 the peak identities are likely to be:

1658: Protein

1444: Phospholipids (phosphatidylserine, cholesterol, sphingomyelin).

828: Proline, hydroxyproline, phosphatidylserine

A main conclusion from comparing average spectra with the literature is therefore that the Raman spectra from my study closely match lab based Raman spectra but do not match the only over *in vivo* study. A possible explanation is that the main difference between spectra taken in the controlled laboratory environment and the operating theatre is the presence of significant amounts of background fluorescence, as shown in section 3.2.2.7 has shown that fluorescence from the tissue itself carries diagnostic information so it is possible that the fluorescence alone contributes to the difference between the spectra obtained in the study by Jermyn *et al.*

A second conclusion is that the protein and phospholipid parts of the Raman spectrum are key elements in distinguishing tumour from normal tissue. This central task of the surgeon is addressed next.

6.3.1.1. Tumour detection

The average normal and tumour spectra from all 1171 *in vivo* spectra showed significant differences (Figure 45). The key peak which was strongly conserved between tumour and normal was the 1444 peak. This can be identified as a phospholipid and cholesterol peak based on comparative *ex vivo* work from Leslie *et al* (Figure 162) and from comparison with lipid reference spectra taken using our system (Figure 45). Bergner et al (Bergner et al., 2012) showed similar peaks in their normal tissue spectra in Figure 163 obtained *ex vivo* using a Raman microscope. The 2 highest intensity peaks are seen in the normal spectra obtained in this thesis (peaks 1439 and 1657 in Figure 163).

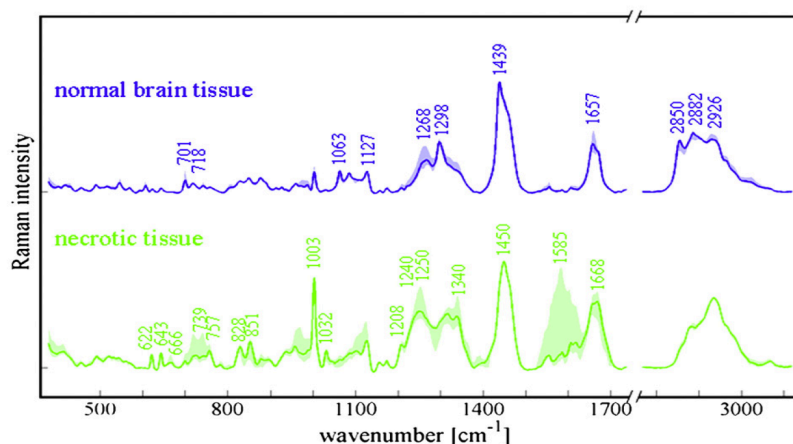


Figure 163. Raman spectra of normal brain tissue and necrotic tissue.

Measured on ex vivo tissue using a Raman microscope. (Bergner et al., 2012)

Nearly all other regions of the Raman spectrum showed significant ($P < 0.05$) differences between normal and tumour spectra. The most noticeable regions were 750-800 (phospholipid) and 1500-1700 corresponding to haemoglobin and protein. Part of the contribution to the higher 1500-1700 wavenumber readings in tumour spectra may be bleeding around the tumour bed, but it should be remembered that gliomas are vascular and require this blood supply to grow aggressively so this may not be completely artefactual.

Extrapolating the spectrum back into the DNA wavenumber region as shown in Figure 46 reveals that DNA also makes an important contribution to the difference between normal and tumour average spectra. It is the assumption that some of these differences are used by the models which discriminate between normal and tumour spectra.

PCA analysis of Normal vs Tumour spectra revealed good separation visually as in Figure 49. It is likely that GBM specimens labelled as 'normal' may have had microscopic disease or may be precancerous as samples were taken close to a focus of high grade cancer in what is known to be a field change disease. It is worth noting, however, excellent performance of the PLS

The best model was formed using RF and SVM reaching 100% sensitivity and 87% specificity with a randomised matched training set (94 normal and 94 tumour spectra). The SVM model required 71 loading vectors indicating a complex optimisation of the tumour/normal boundary. As a Gaussian radial kernel was used and not a simple linear kernel for creating a feature space, relating the loading vectors to wavenumbers was not possible. The RF model used the following peaks as important features:

1628 (close to 1658, Protein)
748 (close to 755, Phospholipid and haemoglobin)
1425 (close to 1444, phospholipids)

The RF model did not use the DNA region represented by lower wavenumbers in its decision making rules.

PLS-DA analysis reached a sensitivity of 92% and specificity of 85% when Raman data was combined with tissue autofluorescence (measured by not subtracting the background fluorescence from the overall signal after stimulation with the 785nm laser). The combination of autofluorescence increased the accuracy of the PLS-DA model. Interestingly, autofluorescence alone was a significant differentiating factor between tissue and normal spectra as shown in Figure 53, reproduced below.

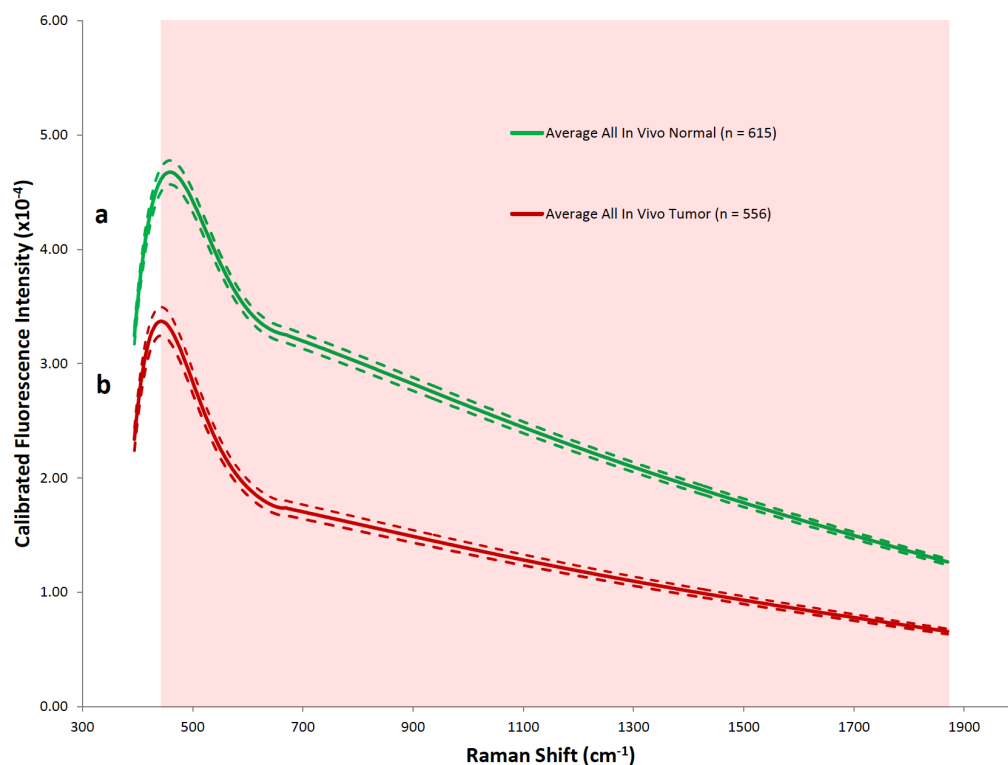


Figure 53. Average auto-fluorescence spectra from brain tissue derived from polynomial fits to the calibrated raw spectra.

From top to bottom the spectra are: a) average of all normal spectra ($n = 615$), and b) average of all tumour spectra ($n = 556$). Dashed lines are the $\pm$ uncertainties on the averages. Shaded wavebands are ranges over which a non-parametric statistic (Mann-Whitney U) indicates a highly significant ($p < 10^{-6}$) difference between the tumour and normal averages.

This correlates well with other studies reporting reduced autofluorescence in glioma tumour tissue compared to normal cortex and white matter (Zanello et al., 2017) (Butte et al., 2010). This signal can therefore be used to augment Raman information and improve accuracy further.

Enlarging the dataset by including *ex vivo* data to build the models resulted in 92% sensitivity and 80% specificity with an AUC of 94% using SVM. This suggests that *ex vivo* data can be used to augment and build larger training datasets for future models. The ROC curve for this model is reproduced below in Figure 65.

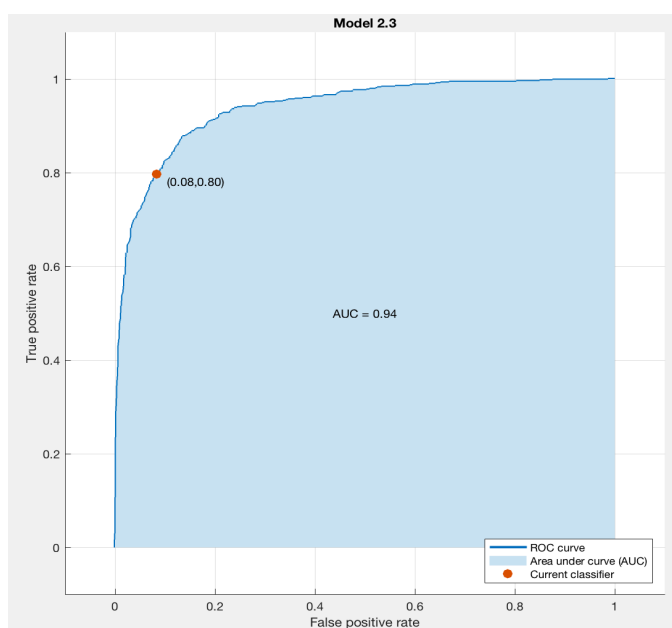


Figure 65. ROC curve of the cubic SVM classifier for 2226 *in vivo* and *ex vivo* spectra.

The AUC is an impressive 0.94 suggesting good performance of this classifier.

6.3.1.2. Tumour density

Tumour density was best modelled with SVM with fluorescence data, and in many ways is an extension of the previous section looking at tumour detection. This produced a model with 83% sensitivity and 82% specificity for distinguishing 3 classifier labels relating to infiltration:

Tumour load	Classifier Label
Zero	Normal
Intermediate	Infiltrated
Solid tumour	Tumour

This finding confirms other work suggesting Raman spectroscopy is able to distinguish tumour density using SRS microscopy (Ji et al., 2015a) and direct *in vivo* spectroscopy (Jermyn et al., 2015a). The aim would be to give the surgeon a more objective indication of tumour density in the tissue being excised when determining the resection margin. This is most important in GBM where the disease is known to be a continuum with tumour cells existing at their highest concentration where the tumour is visualised on MRI but also spreading more distantly even to the opposite 'normal' looking hemisphere on MRI. Being able to objectively identify tumour infiltration in brain tissue would help determine resection in the no man's land beyond MRI contrast enhancement where recurrence from tumour tissue is the most frequent cause of treatment failure and yet where resecting too far results in neurological deficit.

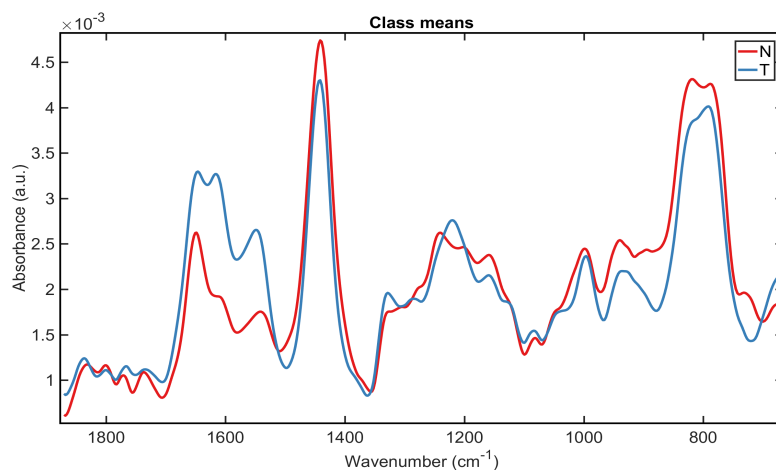
6.3.1.3. Tumour grade

Tumour grades segregated well on PCA and LDA models and on average spectra as shown in Figure 42. In particular grades 2 and 3 separate very well simply on Raman averages as shown in Figure 70. This is significant as differentiating grade 3 from grade 2 tumours can be challenging using routine histopathology. This was the only clinical feature which separated well on all measures prior to model training. Good results were obtained with all machine learning algorithms but the best results were with SVM, Raman and fluorescence combined reaching an accuracy of 98% and a ROC with an AUC of 100%.

This data suggests strong models can be generated to grade tumour tissue using Raman spectroscopy. In contrast to this Jermyn *et al* could not distinguish independent WHO tumour grades accurately and could only distinguish WHO grade 2 from WHO grades 3 and 4 with 82% accuracy as a 2-class classification problem. They could not distinguish between grades 3 and 4. This may in part be due to the lack of features in their over-smoothed spectra. None of the *ex vivo* Raman studies on human brain tumours published looked at WHO grading using spectra making the results from this thesis an important contribution to the literature.

6.3.1.4. Tumour type

Different tumour types were shown to have uniquely different spectra. Figure 164 shows the average spectra of GBM and astrocytoma.



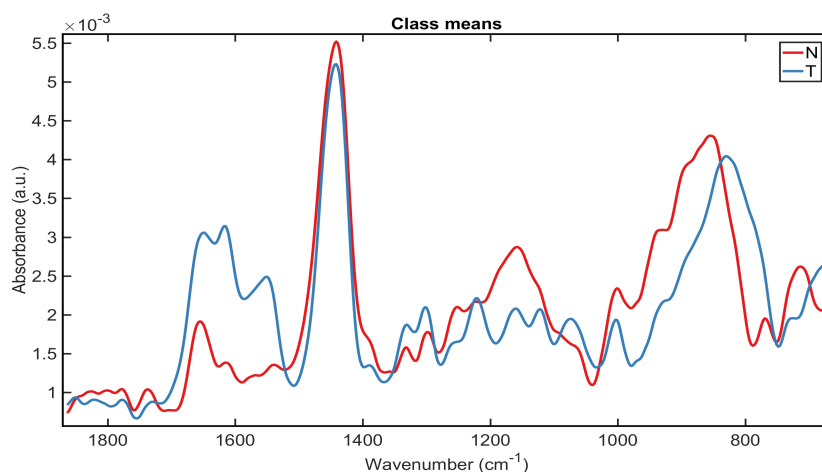


Figure 164. Average *in vivo* Raman spectra of GBM and astrocytoma tissue. Average spectra of GBM (above) and astrocytoma (below) Note: axes have been reversed in this figure.

With GBM the highest intensity peaks include 800, 1220, 1440, and a broad shoulder at 1550-1650. There is some overlap with published *ex vivo* spectra from Kalkanis et al (Figure 165) and *in vivo* spectra from Jermyn et al (Figure 166). Unfortunately, the spectra from Jermyn et al are too noisy to compare peaks accurately, with only the 1440 peak being shared in a spectrum which has a wandering baseline, or overriding fluorescence.

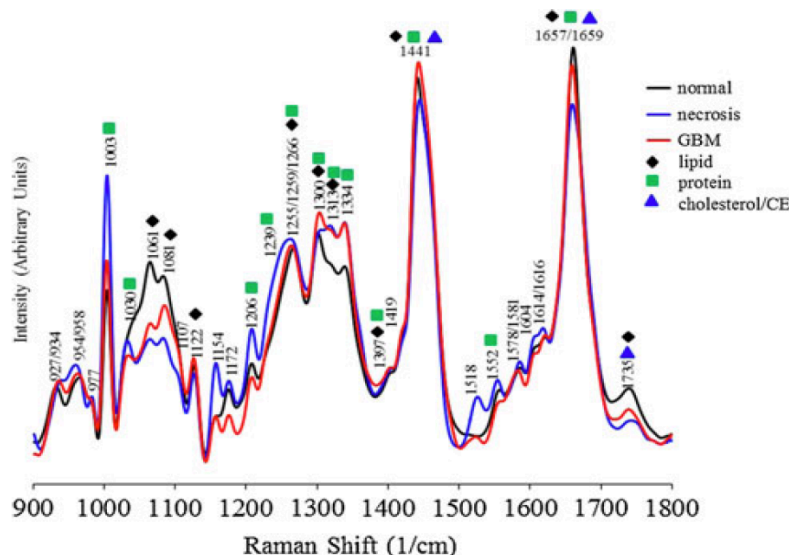


Figure 165. Average *ex vivo* GBM and normal spectra from Kalkanis et al. (Kalkanis et al., 2014)

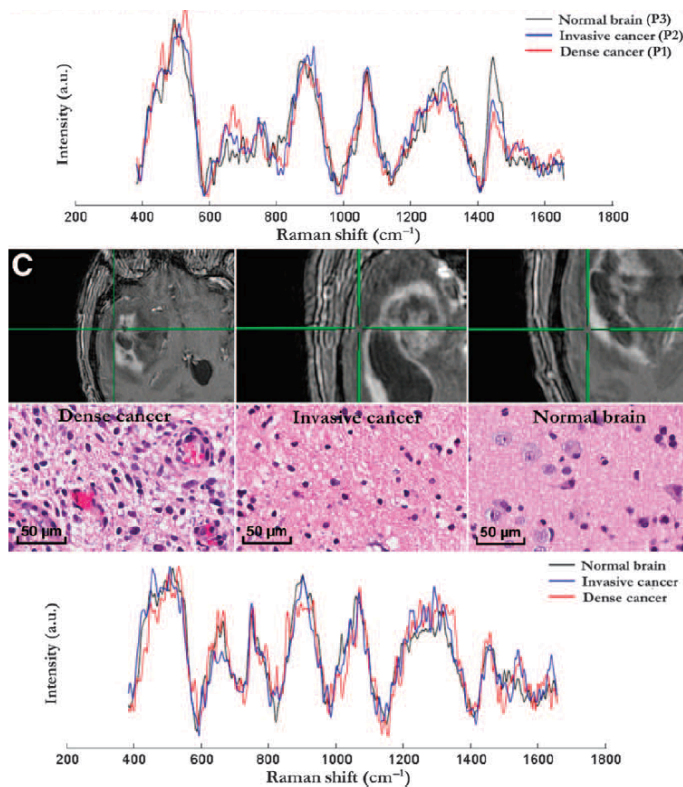


Figure 166. Normal and tumour spectra from single patients from Jermyn et al.
Above: Astrocytoma. Below: GBM (Jermyn et al., 2015a)

Multivariate analysis revealed excellent separation between glioma and non-glioma classes on LDA, and excellent separation on the PCA plot which is shown below in Figure 77.

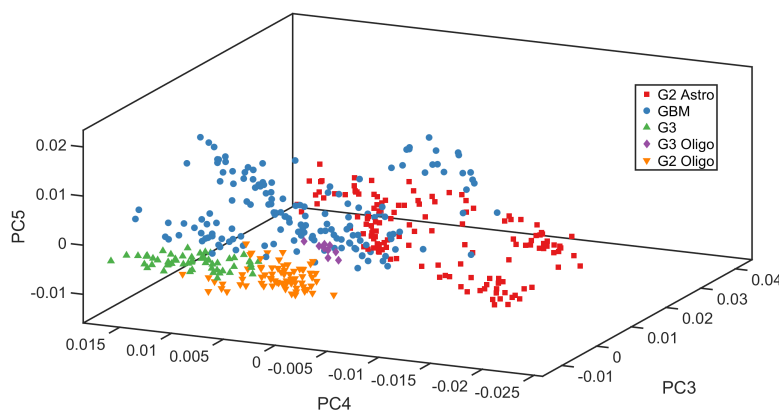


Figure 77. PCA plot showing PCA components 3,4 and 5.
Excellent segregation of spectra into tissue diagnostic groups is shown, with different grades of tumour within the same histological type segregating also.

These results are similar to those reported by Gajjar et al in their *ex vivo* series looking at diagnostic segregation using PCA-LDA models on Raman microscopy in 52 patients (Gajjar et al., 2012b). The tumour type SVM model in this thesis went on to predict tumour type with a high accuracy of 94%, outperforming the PLS-DA

model which had a sensitivity of 85% and specificity of 80%. This is better than the only other published *ex vivo* human glioma type predictive model which produced a best accuracy of 89% for astrocytoma (Leslie et al., 2012).

6.3.1.5. Tumour margin

There has been no data published on margin analysis by the only *in vivo* study on Raman spectroscopy for brain tumours. The data from this thesis is an original contribution to the literature. It shows for the first time *in vivo* spectroscopic data processed by a machine learning algorithm in 3D space overlaid onto the navigational MRI. Figure 81 is reproduced below and shows a snapshot of the 3D model produced.

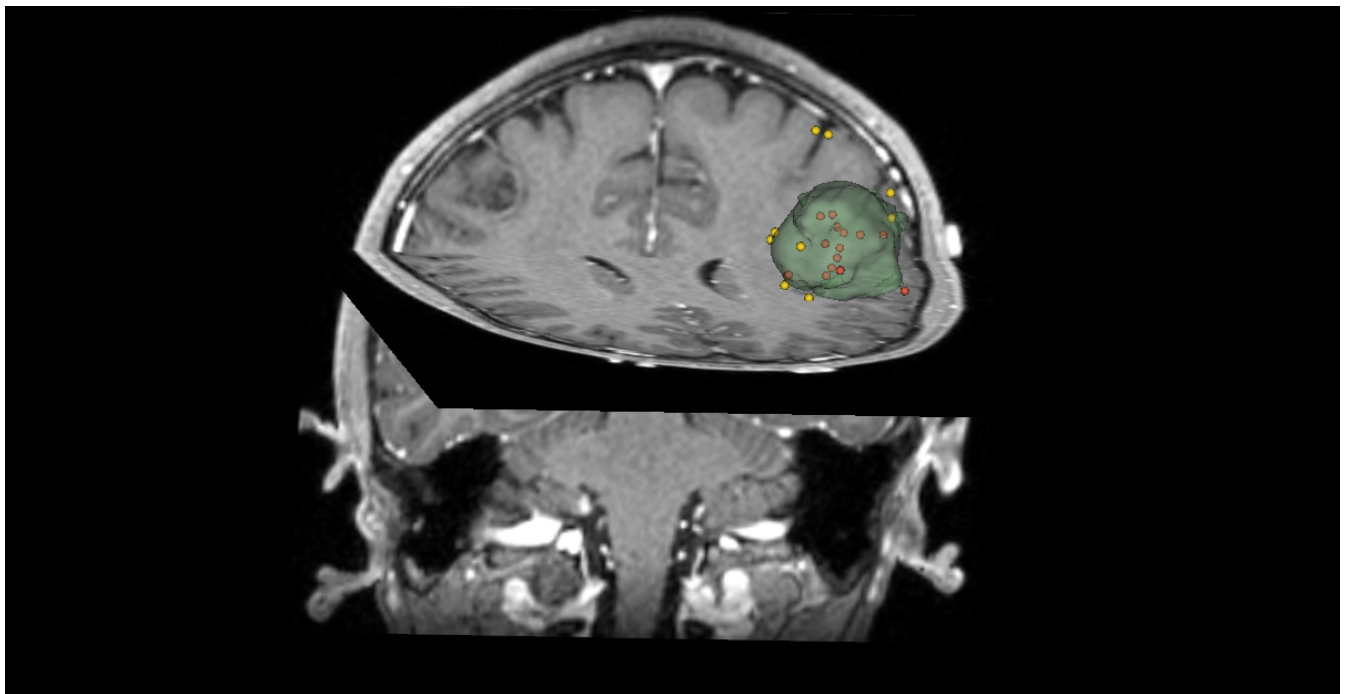


Figure 81. 3D visualisation of a brain tumour showing Raman data analysis points and highlighting the radiological tumour border , A/P view.

The radiological border is the edge of the tumour volume (green) whereas the chemical border is based on Raman sampling points, yellow=normal tissue, red=tumour tissue. All red points lie inside the tumour volume apart from one outlier seen on the right side of the tumour.

The results are striking. The tumour volume based on T2 FLAIR imaging has been outlined in green. The algorithm has correctly identified the tumour and surrounding normal brain with one visible outlier in red. This was identified on histopathology as tumour and was in radiologically 'normal' brain and was therefore correctly identified. Although caution must be exercised as this is only a single case this example does illustrate the power of Raman spectroscopy to guide surgery with stereotactic molecular data.

6.3.1.6. Genetic status

Raman spectroscopy has already been demonstrated *in vivo* as being able to distinguish normal tissue from tumour tissue in the human brain from data presented in this thesis and one previous study (Jermyn et al., 2015a). However, its ability to detect distinct molecular phenotypes has been difficult to demonstrate *ex vivo* (Gajjar et al., 2012a), and there have been no *in vivo* studies to date in part due to the challenges of obtaining adequate *in vivo* Raman spectra. We have shown it is possible to acquire high quality spectra in the clinical environment where ambient background light and signal to noise ratio have previously been a problem. The spectra obtained here were obtained under constant visible illumination and were of high enough quality to accurately predict, with high sensitivity and specificity, the IDH, ATRX and MGMT status of glial tumours. DNA, lipid and protein regions of the Raman spectrum are used by the predictive models we have generated, suggesting that changes at multiple biological scales are responsible for the differences observed in glioma. Peak analysis has revealed some tentative underlying mechanisms. The strong DNA backbone signal seen in ATRX mutated tissue correlates well with the known involvement of ATRX in chromatin remodelling, specifically the binding of the H3.3 histone to telomeric regions of chromosomes (Lewis et al., 2010). Dysfunction of this pathway is thought to lead to genomic instability through the ALT phenotype (Kannan et al., 2012). The key differences seen in MGMT promotor methylated tissue compared to non-methylated tissue are mainly due to the DNA region of the Raman spectrum, which matches the observation that methylation of the MGMT promotor silences MGMT expression and impairs DNA repair causing accumulation of damaged, hypermethylated DNA (Thon et al., 2013). The IDH mutant spectrum correlates well with 2-HG, a metabolite which is known to accumulate in IDH mutated cells (Dang et al., 2009).

This analysis is in real time, takes about 2 seconds, is label and preparation free and readily available within the operative field of the surgeon during the procedure. Used in this way, *in vivo* Raman spectroscopy can identify good and poor prognostic groups of patients immediately, allowing the surgeon to tailor the aggressiveness of surgery appropriately and to select patients for local chemotherapy treatment at the time of the first operation (Vogelbaum and Aghi, 2015). The consequent stratification of patients on the basis of molecular information introduces Raman spectroscopy as an important platform for the introduction of precision medicine to cancer and opens the way for matching Raman profiles with existing and novel treatment modalities, streamlining care and ensuring the right treatment is selected for the right patient.

6.3.2. iKnife: REIMS

REIMS is a unique mass spectrometry system with only one paper available for comparison by Balog et al (Balog et al., 2013). Other studies by Eberlin and Santanaga use DESI which is a near patient mass spectrometry imaging technique which can not currently be used *in vivo*. Direct mass spectrum comparison of averages is therefore not possible. The DESI studies were performed on an extremely small number of patients (5 for Eberlin et al., 2 for Santanaga).

Balog et al. did not present *in vivo* PCA or LDA data but their *ex vivo* LDA data shows in Figure 167 good separation of different tumour types in 37 patients.

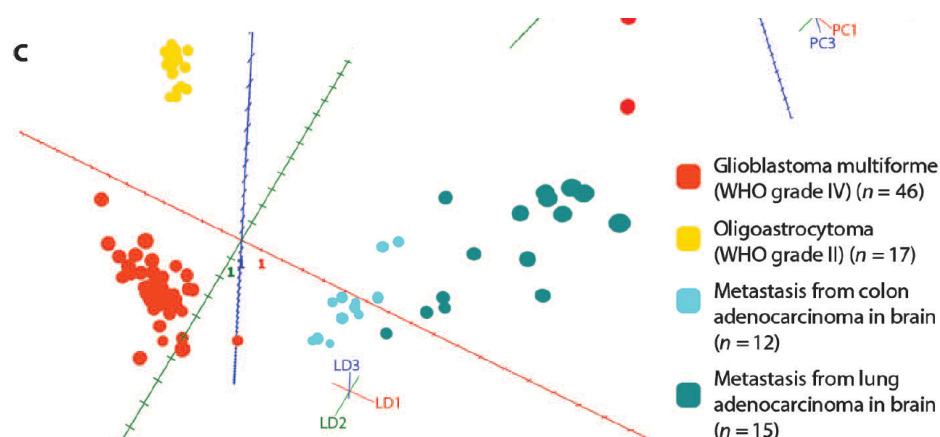


Figure 167. LDA plot of *ex vivo* samples from the first use of the iKnife (Balog et al., 2013)

The series of patients in this thesis represents the largest set of patients with *in vivo* mass spectrometry data from brain tumours (75 patients).

6.3.2.1. Tumour detection

Best results from PCA-LDA modelling of non-IPA data using LOPOCV revealed 83% accuracy for tumour vs normal classification. This is increased slightly by using less robust leave 20% out CV to 86.3% accuracy. Combining IPA and non-IPA datasets from all 75 patients in this study using LOPOCV gives 80.1% accuracy. This is very similar to the results from Balog et al's *in vivo* tumour series with 83.7% diagnosed as 'glioma' vs mets (92.3) vs normal (89.5) in 37 patients from 90 tissue samples.

Eberlin et al achieved 100% tumour tissue sensitivity but this was based on small number of 5 patients and 21 samples and must be interpreted with caution.

6.3.2.2. Tumour density

Highest accuracy was with the PCA-LDA model of 2 tumour density classes (high and low) which yielded LOPOCV accuracy of 88.5% using IPA data. This is significant as DESI is unable to classify tumour density accurately as shown in Table 42, and the previous use of REIMS for a small number of gliomas did not investigate tumour density (Balog et al., 2013).

Name	Diagnosis	Tumor type (%)		Glioma subtype (%)			Glioma grade (%)			Glioma concentration (%)		
		GL	MG	A	OA	O	II	III	IV	Low	Med	High
Case 5												
S28	O, III, med	34	66	12	26	63	0	93	8	4	67	29
S29 (A)	O, III, med	68	32	10	7	83	1	24	75	26	74	0
S30 (C)	O, III, med	84	16	47	0	53	2	33	64	11	78	11
S31 (D)	O, II, low	85	15	46	17	38	0	6	94	48	52	0
S32 (B)	O, III, med	63	38	35	0	65	3	45	53	43	43	15
Total		56	44	22	15	63	1	57	42	19	65	16

Table 42. SVM classifications for DESI on glioma samples for a single patient from Eberlin et al. (Eberlin et al., 2013)

Poor grade and tumour density prediction is shown using SVM on DESI data.

The ability to objectively measure tumour density enables the surgeon to detect infiltrated tissue at the resection margin and fine tune the matching further. Together with Raman spectroscopy this represents one of the only *in vivo* technologies capable of detecting tumour infiltration in real time.

6.3.2.3. Tumour type

The iKnife has been demonstrated in this thesis as being capable of classifying tissue as glioma or non-glioma with 95.3% accuracy. This compares well with the findings of previous use of the iKnife which gave sensitivity varying from 83.7% to 92.3% (Balog et al., 2013).

The iKnife was not demonstrated as being able to classify on the basis of tumour grade. This is likely to be a true effect of the data. All WHO grades were represented in the 75 patients in the study, and whilst other tissue features have been classified accurately WHO grade remains a challenge. WHO grading has not been investigated by any other studies. The poor performance of the models may be due to the fact that WHO grading is an artificial clinicopathological concept the molecular determinants of which are not lipid based.

6.3.2.4. Genetic status

IDH was detected by (Santagata et al., 2014) using DESI MS/MS to show that m/z 146.7 was detected as 2-HG, a marker for IDH mutation. 2-HG was measured as the pure compound and from mouse xenograft models from IDH mutant tumour to confirm m/z 146.6 corresponded with 2-HG rich tissue. No data regarding sensitivity or specificity was published in this small *in vivo* series of 2 patients.

The data presented in this thesis is from a much larger patient cohort. Detection of IDH mutated tissue was modelled successfully using non-IPA data and in GBM tissue using the mass range 100-1000 in order to capture 2-HG. Leave 20% out CV confirmed 95.8% sensitivity for non-IPA data and 94% accuracy for detecting IDH mutation in GBM. LOPOCV results were however disappointing. This may be due to the fact that the numbers of patients being included in the model was so small (18 patients) that leaving one patient out resulted in a failed training set.

Models trained to identify tissue with the ATRX mutation using REIMS on 47 patients performed poorly, with over classification of normal ATRX spectra as mutated. Again, as this was a subset analysis it could be that patient numbers were not high enough to produce a balanced dataset for training the model. The same may be true for the poorly performing MGMT methylation models built from 15 patients.

6.3.3. Conclusion

This chapter has compared Raman and iKnife model results and spectrum features with similar studies in the literature. The diagnostic performance of these 2 novel *in vivo* modalities has been summarised and discussed. A head to head comparison is difficult as both technologies collect data on overlapping but non-identical sets of molecular data. Raman spectroscopy analyses chemical bonds in DNA, lipids and proteins whereas the iKnife analysis phospholipids in great detail.

Table 43 summarises a direct comparison between the 2 technologies.

Diagnostic performance	Raman Spectroscopy		REIMS/iKnife	
	Sensitivity (%)	Specificity (%)	Sensitivity (%)	Specificity (%)
Tumour Detection	100	87	79	89
Tumour Density	83	82	88.5	91.6
Tumour Grade	98	98	61.6	60.6
Tumour Type	94	90	<50%	<50%
Tumour Margin	98	98	n/a	n/a
Genetic Status				
IDH	98	93	71.9	70.3
ATRX	99	96	84.1	66.9
MGMT	96	96	63	55.1
System features	Raman Spectroscopy		REIMS/iKnife	
Tissue destruction	None		Diathermy	
Surgical tool integration	Extra tool		Already integrated	
Speed	1s		2-3 seconds	
Reliability	Excellent		Prone to blockage	
Portability	Small cart		Large trolley	
Noise	Quiet		Noisy vacuum pump	
Laser safety	Required		Not required	

Table 43. Summary table comparing diagnostic performance and system features of Raman spectroscopy versus iKnife.

Blue coloured boxes indicate superiority of one technology over the other.

The question arises as to which technology is best for the surgeon and patient. The clinical utility of real-time tissue characterisation is 2 fold:

1. Diagnostic analysis of tissue: biopsy support
2. Classification of tumour vs normal/ infiltrated tissue: resection margin control

Hence this technology can assist in 2 of the most important aspects of brain tumour surgery: tissue diagnosis and control of the resection margin. Raman spectroscopy offers a non-destructive and faster alternative to the iKnife in both of these use cases. As shown in Table 43 Raman spectroscopy out performs the iKnife in terms of diagnostic accuracy, reliability and portability with the only drawback being the laser safety control requirements for the operating theatre and the fact that the probe represents an extra surgical instrument. In the surgeon's hands Raman spectroscopy seemed more intuitive despite the iKnife being integrated into bipolar forceps.

6.4. Limitations of this thesis

Methodological limitations of system construction and data analysis have been discussed in the relevant chapters. There are some general themes which will be mentioned here. These issues will inform further trials which will follow on from this work and in related areas of machine learning and clinical cancer research.

As with all machine learning studies, the predictive ability of the systems being developed is limited to the data collected and compared in this study. The current focus was understandably on tumour diagnostics and characterisation. The system cannot, for example, currently distinguish between infection or tumour which would also be useful for clinical decision making as cases with infection were not included in the training dataset. For use in the emergency setting (when scans may be rudimentary and histopathology not available) this may be more important. Genus level species differentiation using the iKnife is possible (as detailed in section 2.7.3.5) and so this may be possible with a larger REIMS dataset in the future.

As well as the quality of data the amount of data gathered is a key issue. This directly affects the accuracy of diagnostic algorithms in many ways. The main contribution that a large data source makes is the statistical robustness of models generated from it, as enough data remains in subgroups for detailed analysis to be meaningful. The other aspect is that gathering large sets of data for brain cancer is only possible when multiple centres collaborate and share information. This provides a valuable opportunity of centre to centre variation to be measured and investigated which impacts the accuracy of the system when used in different sites and patient populations. There is controversy on whether machine learning algorithms developed in one population sample are applicable for the whole population if key differences exist. One of the key themes to emerge from this work is importance of balanced datasets for training which contain the clinical features required to create accurate predictive models. It seems for example that to train models to recognise tissue with IDH mutation a sample size of over 400 tissue samples is required (section 3.7.10.2). The poor performance of REIMS models for this classification task can be explained by the fact that these sample sizes were not reached.

A key limitation of the iKnife was the switch from non-IPA to IPA from patient 36 onwards. This rendered the data from the first 36 patients incompatible with the

subsequent patients in the study for the vast majority of data comparisons made because of differences in the mass spectrum caused by adding IPA. Although this improved subsequent spectra valuable data concerning MGMT methylation and IDH mutation from low grade patients operated on earlier in the trial was unavailable for model generation in a larger patient dataset resulting in poor performance in genetics prediction for the iKnife.

The models generated in this study serve as a proof of concept from 75 patients in a single centre. The generalisability of these models to other neurosurgical centres and patient populations remains to be examined as part of a multicentre study which is currently being designed and will be discussed in the next chapter. It is likely that such a feasibility study will involve further data collection to increase the accuracy of the models generated and add new features to the predictions possible such as clinical outcome prognostics and other molecular biomarkers of disease diagnosis, prognosis and treatment response.

A further limitation is to do with cross validation techniques. The majority of papers published using Raman data from human tumours have used LOOCV or k-fold validation which can be susceptible to overfitting from noise (Kalkanis et al., 2014) (Bergner et al., 2012) (Leslie et al., 2012) (Orringer et al., 2017) (Jermyn et al., 2015a). Only one mass spectrometry study used LOPOCV which mitigates this significantly (Balog et al., 2013). LOPOCV may however only be possible with larger datasets as leaving one patients data out completely may result in an invalid training set in small data sets.

Identification of precise molecular species using Raman spectroscopy remains a thorny issue. It is traditionally done by measuring reference chemicals using the Raman system being investigated and then matching new information with respect to the database of reference chemicals generated *in situ*. It is our hope that our current database of 29 patients will contribute to such a database of reference pathologies. Until a comprehensive database is established containing all the variation seen in clinical practice. Identifying individual chemicals responsible for the differences observed is challenging. Comparison with already published databases and tables is possible but there is no international standard for calibrating Raman measurements to ensure direct comparison of information, with different systems having slightly different peak positions for similar compounds. What doesn't change is the overall spectrum profile and peak to peak ratios which are being measured in this study. REIMS suffers from a similar problem, although with large volume tissue samples MS/MS can be used to repeatedly perform mass spectrometry on an m/z range to fragment molecule clusters further to a point where individual species can be identified. Such large volumes of brain tissue can be difficult to collect. The number of samples possible in one operation averaged at 6 samples per case in this study. In my experience it was too time consuming to obtain more samples with the risk of bleeding with each biopsy and the risks to the patient of prolonging general anaesthetic whilst more samples were being obtained.

A related issue to the rarity of brain tissue is the difficulty in obtaining normal brain tissue as controls. It is unethical to know sample healthy normal brain tissue when it is not part of routine surgery. In some cases, a minimal volume of normal tissue has to be entered where it is safe to do so in order to access a tumour. In other rare

cases tissue of uncertain appearance may be biopsied and may be reported as normal tissue. The volume of normal tissue is therefore very small when compared to the volume of tumour tissue being removed and this can bias learning algorithms towards overclassifying new data as tumour tissue. The training groups can be matched to mitigate for this but it is not a substitute for having a larger group of normal tissue. This ethical challenge affects all machine learning approaches to normal versus tumour tissue classification. Possible work arounds are to collect normal tissue from patients without brain tumours (for example, epilepsy surgery patients) but all patients having surgery have a pathological condition which will bias the data and models generated in some way.

A key limitation of the use of the iKnife is the fact that this technology destroys tissue. By definition the volume of tissue which has generated the mass spectrum has been ionised and destroyed meaning that only histological evaluation of closely neighbouring areas of tissue is possible. This is a problem for highly heterogeneous tumours such as GBM. This can only be mitigated by having as large a tissue database as possible to take into account sample to sample variation from this effect. This is not an issue for Raman spectroscopy as this is a non-destructive technique.

6.5. Future Trials and developments

The data in this thesis shows that Raman spectroscopy and REIMS can be used during surgery to generate molecular information which can be used in real time to profile tissue as tumour or normal and detect tumour density. Raman spectroscopy can go further and accurately predict tumour type, WHO grade status of tissue and can identify tissue containing key genomic alterations seen in gliomas. In this way Raman spectroscopy can act as a platform for bringing a precision medicine approach to neurosurgery. REIMS requires further data and development to reach this stage but in time both technologies can be strong platforms for intraoperative molecular analysis. This chapter discusses further developments for both of these platforms and plans for a multicentre Raman feasibility study.

6.5.1. Future development of Raman

Like any modern technology Raman systems can be divided up into hardware and software. Both elements will be discussed.

6.5.1.1. Hardware

Our current system consists of a single 785nm laser connected with advanced optical fibres and filters to a CCD sensor. Through the use of this system in over 29 cases during neurosurgery it has become apparent that a number of improvements

can be made to this system to improve the data collected with the drawback of adding cost to the system.

Starting with the laser it may be possible to add a second or third laser which is not in the near infrared range thus creating a system not susceptible to IR signal from a range of sources in the operating theatre. There is a suggestion that the higher wave numbers which would become available carry diagnostic information (Ji et al., 2015b) but such techniques currently use a microscope stage as part of an SRS approach. Incorporating this as a fibre optical probe could bring this approach directly to the surgical operating field. Having larger fibres could also increase the amount of data collected as long as it does not affect the surgical ergonomics of using the probe.

Electron Multiplying CCD's or EMCCD's offer faster detection resulting in high resolution Raman image maps and the possibility of quantification of Raman signal. This could allow more detailed, quantitative analysis.

A variety of advanced techniques such as Spatially Offset Raman Spectroscopy (SORS) and Surface Enhanced Raman Spectroscopy (SERS) exist and need to be adapted for *in vivo* use. SORS offers true non-invasive analysis without having to make a surgical incision, and SERS offers very high signal to noise ratio. However, significant obstacles exist with regards to the need for a scattering material such as calcium for SORS and the need for a method of treating the surface of the material being analysed for SERS.

Furthermore, Raman systems can be embedded in catheters and tubes currently widely used in neurosurgery to allow continuous sensing of tissue and fluid in External Ventricular Drains and brain tissue sensors such as intracranial pressure bolts. Such systems can also be used for detection and quantification of therapeutic chemicals crossing the blood brain barrier into brain tissue for which data is lacking.

6.5.1.2. Software

The main contribution to advancing future *in vivo* Raman spectroscopy is in model generation and testing and in representing the model output to the surgeon.

Models can be improved by collecting more detailed data relating to tumour density to enable a quantitative objective measure of tumour infiltration. This may also lead to a higher accuracy of tumour detection by changing the binary tumour classification task (normal versus tumour) into a quantitative spectrum which more realistically represents tumour cell distribution *in vivo*. This would add to surgical utility by helping the surgeon define their own resection margin objectively.

Representing the outputs of the improved models to the surgeon is challenging. There is the danger of creating a sea of information and confusion when the surgeon requires a targeted goal based approach to progress through a complex operation quickly and safely for the patient. One solution is for the output to be sonified, with auditory tones for tumour and normal tissue. Representing this as a quantitative

scale is a challenge. Another option is for data to be relayed back to the guidance systems already in use by the surgeon such as the navigation system as already presented in this thesis, or even as a virtual display on the operating microscope. virtual displays have not been adopted widely in neurosurgery as they are currently either too obtrusive or not noticeable enough for the surgeon to use the right information at the right time during surgery. A compromise may therefore be a Raman output connected to neuronavigation and with an audible sound for certain thresholds.

6.5.2. Future development of iKnife

The iKnife is a complex piece of engineering: a system capable of detailed chemical analysis used in the laboratory. The challenge has been in miniaturising the hardware and enabling it to be used quickly during surgery.

The current version of the iKnife measures a wide mass range from 100 to 1000 (and beyond in some cases). Only a small portion of this range has been shown in this thesis as carrying diagnostically valuable information (mainly the height 600-1000 range and in some cases the lower range of 100-300). It is therefore conceivable that a slim-down version for use in neurosurgery can be created which is smaller, more portable and gives faster response times by quickly detecting only relevant mass ranges.

Further work on maintaining a quieter vacuum would be welcomed by the surgeons and operating theatre staff. The noise pollution is significant at the end of a 4 hour operation with multiple members of staff asking for the machine to be switched off as early as possible for this reason.

The bipolar forceps used with the iKnife have been key to its easy integration with surgical workflow as this does not represent a new instrument for the surgeon to get used to. The irrigation channels in the forceps did however block regularly during surgery resulting in data loss when the surgeon did not have time to unblock the system at a crucial moment in surgery. Replacing the tissue or patient interface of the machine with another modality is currently being investigated. One possibility is the use of a low powered laser to create tissue ionisation with minimal heat effect and tissue damage. This would circumvent the issues with regards to accurate histopathology reporting on tissue as significantly less tissue is destroyed using this technique. The only obstacle to using this system is the need for regulatory and safety approval as ablative lasers are currently not used in neurosurgery.

6.5.3. UK multicentre feasibility study

At a recent meeting of the NIHR Efficiency and Mechanism Evaluation (EME) subgroup within the Society of British Neurological Surgeons in May 2018 it was decided that there was a need for a multicentre feasibility study for Raman spectroscopy based on the successful pilot study data available so far. It was

recognised that further work needed to be done to investigate system performance multiple sites and establish system accuracy prior to launching a randomised controlled trial of surgical intervention based on Raman data to see if patient care could be improved. It was felt the iKnife required more development before it could reach this stage.

The main research questions during the feasibility study will be:

- Can surgeon held Raman Spectroscopy used during surgery increase the amount of tumour tissue safely removed in patients undergoing craniotomy for brain tumours compared to standard surgery?

This can be further split into 2 questions.

- Does Raman Spectroscopy have >95% sensitivity and specificity for identifying tumour tissue compared to histopathology as a gold standard? This is answered by the diagnostic accuracy component of the study (a feasibility study across multiple sites).
- Does the subsequent use of Raman Spectroscopy by the surgeon during surgery lead to an increased volume of tumour tissue being safely removed? This will be answered by the diagnostic utility component of the study (the first multicentre Randomised Controlled Trial ever done in this area).

Our primary hypothesis based on our pilot data in this thesis is that Raman spectroscopy is capable of accurately differentiating normal healthy brain tissue and tumour tissue *in vivo* using histological examination as the 'gold standard'. Our secondary hypothesis is that by doing this rapidly and repeatedly during surgery a surgeon will be able to identify a larger volume of tumour to be removed safely.

In order to answer the research question the multicentre accuracy of Raman spectroscopy must first be established before using it as part of a trial during surgery. Our aims, therefore, are to

- Confirm handheld Raman Spectroscopy is a fast, reliable and accurate platform for tumour analysis during brain surgery for the surgeon.
- Confirm that using data from Raman Spectroscopy as an adjunct to standard care a surgeon can better identify tissue to enable safe resection of larger volumes of tumour than currently possible.

The proposed design is a multicentre trial consisting of 2 stages with a stop/go assessment controlling progression from stage 1 to stage 2. Stage 1 consists of a feasibility study to establish data variation between sites and machines, ease of use of the probe during surgery, and diagnostic accuracy of the probe in identifying tumour tissue. Stage 2 consists of a diagnostic utility study in the form of an RCT looking at *in vivo* Raman spectroscopy + standard treatment vs current standard of care in all centres.

With the collection of large volumes of multicentre data, it is hoped that further data and system standardisation can be achieved. Larger training sets of feature rich clinical information will allow a number of diagnostic, prognostic and predictive models to be investigated to fully explore the potential of Raman spectroscopy to improve patient care.

6.5.4. Conclusions

Advances in Neurosurgery as a speciality have been directly linked to imaging technology. CT and MRI scanning have helped to gain access to this organ which is locked away in its own vault and is physically inaccessible without the skills of a neurosurgeon. The advent of the operating microscope and neuronavigation have transformed the speciality from macroscopic to microscopic precision. The problem of tissue detection during tumour surgery is complex. As described in this thesis holistic molecular analysis at the point of care and the use of powerful machine learning methods to find deep associations with clinically useful features of patient care is a powerful model and is the essence of precision medicine.

The transformative power of this new philosophy of precision medicine as an approach to complex, unsolved problems is palpable. There has been a surge in related work in the field in the past year. The significance of the work in this thesis for neurosurgery and neuro-oncology has been recognised nationally by the NIHR when I was invited to talk about new technology trials in research at the Royal College of Surgeons of England and internationally by several awards, including the European Association of Neurosurgical Societies best poster award 2015 and the European Association of Neuro-Oncology award for best oral presentation in 2016.

It is my hope that pioneers brave enough to step outside of their own fields for the sake of alleviating the suffering of others will continue to create new advances from the overlap of the mature sciences in order to create hope for this devastating disease.

References

- NIST: Basic Atomic Spectroscopic Data - Element Names [Online]. [Accessed].
- ABEL, K., DESCHMERTZING, H. & PETERSON, J. I. 1963. CLASSIFICATION OF MICROORGANISMS BY ANALYSIS OF CHEMICAL COMPOSITION I. : Feasibility of Utilizing Gas Chromatography. *J. Bacteriol.*, 85, 1039-1044.
- AKBARI, H., MACSYZYN, L., DA, X., BILELLO, M., WOLF, R. L., MARTINEZ-LAGE, M., BIROS, G., ALONSO-BASANTA, M., O'ROURKE, D. M. & DAVATZIKOS, C. 2016. Imaging Surrogates of Infiltration Obtained Via Multiparametric Imaging Pattern Analysis Predict Subsequent Location of Recurrence of Glioblastoma. *Neurosurgery*, 78, 572-80.
- ALDAPE, K., AMIN, S. B., ASHLEY, D. M., BARNHOLTZ-SLOAN, J. S., BATES, A. J., BEROUKHIM, R., BOCK, C., BRAT, D. J., CLAUS, E. B., COSTELLO, J. F., DE GROOT, J. F., FINOCCHIARO, G., FRENCH, P. J., GAN, H. K., GRIFFITH, B., HEROLD-MENDE, C. C., HORBINSKI, C., IAVARONE, A., KALKANIS, S. N., KARABATSOU, K., KIM, H., KOUWENHOVEN, M. C. M., MCDONALD, K. L., MILETIC, H., NAM, D.-H., NG, H. K., NICLOU, S. P., NOUSHMEHR, H., ORMOND, R., POISSON, L. M., REIFENBERGER, G., RONCAROLI, F., SA, J. K., SILLEVIS SMITT, P. A. E., SMITS, M., SOUZA, C. F., TABATABAI, G., VAN MEIR, E. G., VERHAAK, R. G. W., WATTS, C., WESSELING, P., WOEHRE, A., YUNG, W. K. A., JUNGK, C., HAU, A.-C., VAN DYCK, E., WESTERMAN, B. A., YIN, J., ABIOLA, O., ZEPS, N., GRIMMOND, S., BUCKLAND, M., KHASRAW, M., SULMAN, E. P., MUSCAT, A. M., STEAD, L. & CONSORTIUM, T. G. 2018. Glioma Through the Looking GLASS: Molecular Evolution of Diffuse Gliomas and the Glioma Longitudinal AnalySiS Consortium. *Neuro-Oncology*.
- ALDERTON, G. K. 2011. Tumorigenesis: the origins of glioma. *Nat Rev Cancer*, 11, 627.
- ALEXANDER, J., GILDEA, L., BALOG, J., SPELLER, A., MCKENZIE, J., MUIRHEAD, L., SCOTT, A., KONTOVOUNISIOS, C., RASHEED, S., TEARE, J., HOARE, J., VESELKOV, K., GOLDIN, R., TEKKIS, P., DARZI, A., NICHOLSON, J., KINROSS, J. & TAKATS, Z. 2017. A novel methodology for in vivo endoscopic phenotyping of colorectal cancer based on real-time analysis of the mucosal lipidome: a prospective observational study of the iKnife. *Surg Endosc*, 31, 1361-1370.
- AMERICA, L. I. O. 2014. *ANSI Z136.1 (2014) - Safe Use of Lasers*, Laser Institute of America.
- AMHARREF, N., BELJEBBAR, A., DUKIC, S., VENETO, L., SCHNEIDER, L., PLUOT, M. & MANFAIT, M. 2007. Discriminating healthy from tumor and necrosis tissue in rat brain tissue samples by Raman spectral imaging. *Biochim Biophys Acta*, 1768, 2605-15.
- AZIMI, P., SHAHZADI, S. & SADEGHI, S. 2015. Use of artificial neural networks to predict the probability of developing new cerebral metastases after radiosurgery alone. *J Neurosurg Sci*.
- BAKER, M. J., BYRNE, H. J., CHALMERS, J., GARDNER, P., GOODACRE, R., HENDERSON, A., KAZARIAN, S. G., MARTIN, F. L., MOGER, J., STONE, N. & SULÉ-SUSO, J. 2018. Clinical applications of infrared and Raman spectroscopy: state of play and future challenges. *The Analyst*, 143, 1735-1757.
- BALOG, J., KUMAR, S., ALEXANDER, J., GOLF, O., HUANG, J., WIGGINS, T., ABBASSI-GHADI, N., ENYEDI, A., KACSKA, S., KINROSS, J., HANNA, G. B., NICHOLSON, J. K. & TAKATS, Z. 2015. In vivo endoscopic tissue identification by rapid evaporative ionization mass spectrometry (REIMS). *Angew Chem Int Ed Engl*, 54, 11059-62.
- BALOG, J., SASI-SZABÓ, L., KINROSS, J., LEWIS, M. R., MUIRHEAD, L. J., VESELKOV, K., MIRNEZAMI, R., DEZSÓ, B., DAMJANOVICH, L., DARZI, A., NICHOLSON, J. K. & TAKÁTS, Z. 2013. Intraoperative Tissue Identification Using Rapid Evaporative Ionization Mass Spectrometry. *Science Translational Medicine*, 5, 194ra93-194ra93.
- BALOG, J., SZANISZLO, T., SCHAEFER, K.-C., DENES, J., LOPATA, A., GODORHAZY, L., SZALAY, D., BALOGH, L., SASI-SZABO, L., TOTH, M. & TAKATS, Z. 2010. Identification of Biological Tissues by Rapid Evaporative Ionization Mass Spectrometry. *Analytical Chemistry*, 82, 7343-7350.
- BARONE, D. G., LAWRIE, T. A. & HART, M. G. 2014. Image guided surgery for the resection of brain tumours. *Cochrane Database of Systematic Reviews*.
- BELJEBBAR, A., DUKIC, S., AMHARREF, N. & MANFAIT, M. 2010. Ex vivo and in vivo diagnosis of C6 glioblastoma development by Raman spectroscopy coupled to a microprobe. *Anal Bioanal Chem*, 398, 477-87.
- BENNETT, A. H. 1885. Case of Cerebral Tumour. *Med Chir Trans*, 68, 243-75.
- BERGER, M. S., DELIGANIS, A. V., DOBBINS, J. & KELES, G. E. 1994. The effect of extent of resection on recurrence in patients with low grade cerebral hemisphere gliomas. *Cancer*, 74, 1784-1791.
- BERGNER, N., BOCKLITZ, T., ROMEIKE, B. F. M., REICHART, R., KALFF, R., KRAFFT, C. & POPP, J. 2012. Identification of primary tumors of brain metastases by Raman imaging and support vector machines. *Chemometrics and Intelligent Laboratory Systems*, 117, 224-232.
- BOLT, F., CAMERON, S. J., KARANCSI, T., SIMON, D., SCHAFFER, R., RICKARDS, T., HARDIMAN, K., BURKE, A.,

- BODAI, Z., PERDONES-MONTERO, A., REBEC, M., BALOG, J. & TAKATS, Z. 2016. Automated High-Throughput Identification and Characterization of Clinically Important Bacteria and Fungi using Rapid Evaporative Ionization Mass Spectrometry. *Analytical Chemistry*, 88, 9419-9426.
- BOURBEAU, P. P. & LEDEBOER, N. A. 2013. Automation in clinical microbiology. *J. Clin. Microbiol.*, 51, 1658-1665.
- BROWN, T. J., BRENNAN, M. C., LI, M., CHURCH, E. W., BRANDMEIR, N. J., RAKSZAWSKI, K. L., PATEL, A. S., RIZK, E. B., SUKI, D., SAWAYA, R. & GLANTZ, M. 2016. Association of the Extent of Resection With Survival in Glioblastoma: A Systematic Review and Meta-analysis. *JAMA Oncology*, 2, 1460-1469.
- BUTTE, P. V., FANG, Q., JO, J. A., YONG, W. H., PIKUL, B. K., BLACK, K. L. & MARCU, L. 2010. Intraoperative delineation of primary brain tumors using time-resolved fluorescence spectroscopy. *Journal of biomedical optics*, 15, 027008.
- CAMERON, S. J., BOLT, F., PERDONES-MONTERO, A., RICKARDS, T., HARDIMAN, K., ABDOLRASOULI, A., BURKE, A., BODAI, Z., KARANCSI, T., SIMON, D., SCHAFFER, R., REBEC, M., BALOG, J. & TAKATS, Z. 2016. Rapid Evaporative Ionisation Mass Spectrometry (REIMS) Provides Accurate Direct from Culture Species Identification within the Genus *Candida*. *Scientific Reports*, 6, 36788.
- CANCER GENOME ATLAS RESEARCH, N., BRAT, D. J., VERHAAK, R. G., ALDAPE, K. D., YUNG, W. K., SALAMA, S. R., COOPER, L. A., RHEINBAY, E., MILLER, C. R., VITUCCI, M., MOROZOVA, O., ROBERTSON, A. G., NOUSHMEHR, H., LAIRD, P. W., CHERNIACK, A. D., AKBANI, R., HUSE, J. T., CIRIELLO, G., POISSON, L. M., BARNHOLTZ-SLOAN, J. S., BERGER, M. S., BRENNAN, C., COLEN, R. R., COLMAN, H., FLANDERS, A. E., GIANNINI, C., GRIFFORD, M., IAVARONE, A., JAIN, R., JOSEPH, I., KIM, J., KASAIAN, K., MIKKELSEN, T., MURRAY, B. A., O'NEILL, B. P., PACHTER, L., PARSONS, D. W., SOUGNEZ, C., SULMAN, E. P., VANDENBERG, S. R., VAN MEIR, E. G., VON DEIMLING, A., ZHANG, H., CRAIN, D., LAU, K., MALLERY, D., MORRIS, S., PAULAUSKIS, J., PENNY, R., SHELTON, T., SHERMAN, M., YENA, P., BLACK, A., BOWEN, J., DICOSTANZO, K., GASTIER-FOSTER, J., LERAAS, K. M., LICHTENBERG, T. M., PIERSON, C. R., RAMIREZ, N. C., TAYLOR, C., WEAVER, S., WISE, L., ZMUDA, E., DAVIDSEN, T., DEMCHOK, J. A., ELEY, G., FERGUSON, M. L., HUTTER, C. M., MILLS SHAW, K. R., OZENBERGER, B. A., SHETH, M., SOFIA, H. J., TARNUZZER, R., WANG, Z., YANG, L., ZENKLUSEN, J. C., AYALA, B., BABOUD, J., CHUDAMANI, S., JENSEN, M. A., LIU, J., PIHL, T., RAMAN, R., WAN, Y., WU, Y., ALLY, A., AUMAN, J. T., BALASUNDARAM, M., BALU, S., BAYLIN, S. B., BEROUKHIM, R., BOOTWALLA, M. S., BOWLBY, R., BRISTOW, C. A., BROOKS, D., BUTTERFIELD, Y., CARLSEN, R., CARTER, S., CHIN, L., et al. 2015. Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. *N Engl J Med*, 372, 2481-98.
- CASSAGNE, C., CELLA, A.-L., SUCHON, P., NORMAND, A.-C., RANQUE, S. & PIARROUX, R. 2013. Evaluation of four pretreatment procedures for MALDI-TOF MS yeast identification in the routine clinical laboratory. *Med. Mycol.*, 51, 371-377.
- COBURGER, J., ENGELKE, J., SCHEUERLE, A., THAL, D. R., HLAVAC, M., WIRTZ, C. R. & KONIG, R. 2014. Tumor detection with 5-aminolevulinic acid fluorescence and Gd-DTPA-enhanced intraoperative MRI at the border of contrast-enhancing lesions: a prospective study based on histopathological assessment. *Neurosurgical focus*, 36, E3.
- CROXATTO, A., PROD'HOM, G. & GREUB, G. 2012. Applications of MALDI-TOF mass spectrometry in clinical diagnostic microbiology. *FEMS microbiology reviews*, 36, 380-407.
- DAMADIAN, R. 1971. Tumor detection by nuclear magnetic resonance. *Science*, 171, 1151-3.
- DANDY, W., E 1928. Removal of right cerebral hemisphere for certain tumors with hemiplegia. *JAMA*, 90, 823-825.
- DANG, L., WHITE, D. W., GROSS, S., BENNETT, B. D., BITTINGER, M. A., DRIGGERS, E. M., FANTIN, V. R., JANG, H. G., JIN, S., KEENAN, M. C., MARKS, K. M., PRINS, R. M., WARD, P. S., YEN, K. E., LIAU, L. M., RABINOWITZ, J. D., CANTLEY, L. C., THOMPSON, C. B., VANDER HEIDEN, M. G. & SU, S. M. 2009. Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature*, 462, 739.
- DEPARTMENT OF HEALTH AND SOCIAL CARE, T. A. F. W. G. O. B. T. R. 2018. *Report of the Task and Finish Working Group on Brain Tumour Research*, London, UK, Great Britian Stationary Office.
- DISEASE, N. R. C. U. C. O. A. F. F. D. A. N. T. O. 2011. *Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease*, Washington (DC), National Academies Press (US).
- DRYSDALE, C. M., MCGRAW, D. W., STACK, C. B., STEPHENS, J. C., JUDSON, R. S., NANDABALAN, K., ARNOLD, K., RUANO, G. & LIGGETT, S. B. 2000. Complex promoter and coding region β 2-adrenergic receptor haplotypes alter receptor expression and predict in vivo responsiveness. *Proceedings of the National Academy of Sciences*, 97, 10483-10488.
- EBERLIN, L. S., NORTON, I., DILL, A. L., GOLBY, A. J., LIGON, K. L., SANTAGATA, S., COOKS, R. G. & AGAR, N. Y. R.

2012. Classifying human brain tumors by lipid imaging with mass spectrometry. *Cancer research*, 72, 645-54.
- EBERLIN, L. S., NORTON, I., ORRINGER, D., DUNN, I. F., LIU, X., IDE, J. L., JARMUSCH, A. K., LIGON, K. L., JOLESZ, F. A., GOLBY, A. J., SANTAGATA, S., AGAR, N. Y. R. & COOKS, R. G. 2013. Ambient mass spectrometry for the intraoperative molecular diagnosis of human brain tumors. *Proceedings of the National Academy of Sciences of the United States of America*, 110, 1611-6.
- ELDER, J. B. & RAO, G. 2017. *Intraoperative Imaging, An Issue of Neurosurgery Clinics of North America*, E-Book, Elsevier Health Sciences.
- ELLINGSON, B. M., ABREY, L. E., NELSON, S. J., KAUFMANN, T. J., GARCIA, J., CHINOT, O., SARAN, F., NISHIKAWA, R., HENRIKSSON, R., MASON, W. P., WICK, W., BUTOWSKI, N., LIGON, K. L., GERSTNER, E. R., COLMAN, H., DE GROOT, J., CHANG, S., MELLINGHOFF, I., YOUNG, R. J., ALEXANDER, B. M., COLEN, R., TAYLOR, J. W., ARRILLAGA-ROMANY, I., MEHTA, A., HUANG, R. Y., POPE, W. B., REARDON, D., BATCHELOR, T., PRADOS, M., GALANIS, E., WEN, P. Y. & CLOUGHESY, T. F. 2018. Validation of postoperative residual contrast-enhancing tumor volume as an independent prognostic factor for overall survival in newly diagnosed glioblastoma. *Neuro Oncol*, 20, 1240-1250.
- ELLIS, D. I., COWCHER, D. P., ASHTON, L., O'HAGAN, S. & GOODACRE, R. 2013. Illuminating disease and enlightening biomedicine: Raman spectroscopy as a diagnostic tool. *Analyst*, 138, 3871-3884.
- EMBLEM, K. E., PINHO, M. C., ZOLLNER, F. G., DUE-TONNESSEN, P., HALD, J. K., SCHAD, L. R., MELING, T. R., RAPALINO, O. & BJORNERUD, A. 2015. A generic support vector machine model for preoperative glioma survival associations. *Radiology*, 275, 228-34.
- EYUPOGLU, I. Y., BUCHFELDER, M. & SAVASKAN, N. E. 2013. Surgical resection of malignant gliomas-role in optimizing patient outcome. *Nat Rev Neurol*, 9, 141-151.
- GAJJAR, K., HEPPENSTALL, L. D., PANG, W., ASHTON, K. M., TREVISAN, J., PATEL, II, LLABJANI, V., STRINGFELLOW, H. F., MARTIN-HIRSCH, P. L., DAWSON, T. & MARTIN, F. L. 2012a. Diagnostic segregation of human brain tumours using Fourier-transform infrared and/or Raman spectroscopy coupled with discriminant analysis. *Anal Methods*, 5, 89-102.
- GAJJAR, K., HEPPENSTALL, L. D., PANG, W., ASHTON, K. M., TREVISAN, J., PATEL, I. I., LLABJANI, V., STRINGFELLOW, H. F., MARTIN-HIRSCH, P. L., DAWSON, T. & MARTIN, F. L. 2012b. Diagnostic segregation of human brain tumours using Fourier-transform infrared and/or Raman spectroscopy coupled with discriminant analysis. *Analytical methods : advancing methods and applications*, 5, 89-102.
- GOTTLIEB, S. 2000. Personalised medicine comes a step closer for asthma. *BMJ*, 321, 724.
- GRECH-SOLLARS, M., VAQAS, B., THOMPSON, G., BARWICK, T., HONEYFIELD, L., O'NEILL, K. & WALDMAN, A. D. 2016. An MRS- and PET-guided biopsy tool for intraoperative neuronavigational systems. *Journal of Neurosurgery*, 1-7.
- GREENBLATT, S. H. 2003. Harvey Cushing's paradigmatic contribution to neurosurgery and the evolution of his thoughts about specialization. *Bull Hist Med*, 77, 789-822.
- GUPTA, S. 2015. Trouble in Paris (classification): polyp morphology is in the eye of the beholder. *Am J Gastroenterol*, 110, 188-91.
- GUZE, K., SHORT, M., SONIS, S., KARIMBUX, N., CHAN, J. & ZENG, H. 2009. Parameters defining the potential applicability of Raman spectroscopy as a diagnostic tool for oral disease. *Journal of Biomedical Optics*, 14, 014016.
- HANLEY, J. A. & MCNEIL, B. J. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143, 29-36.
- HAUSER, S. B., KOCKRO, R. A., ACTOR, B., SARNTHEIN, J. & BERNAYS, R. L. 2016. Combining 5-aminolevulinic acid fluorescence and intraoperative magnetic resonance imaging in glioblastoma surgery: A histology-based evaluation. *Clinical Neurosurgery*, 78, 475-483.
- HEFTI, M., VON CAMPE, G., MOSCHOPULOS, M., SIEGNER, A., LOOSER, H. & LANDOLT, H. 2008. 5-aminolevulinic acid induced protoporphyrin IX fluorescence in high-grade glioma surgery: a one-year experience at a single institution. *Swiss medical weekly*, 138, 180-5.
- HERVEY-JUMPER, S. L. & BERGER, M. S. 2016. Maximizing safe resection of low- and high-grade glioma. *Journal of Neuro-Oncology*, 130, 269-282.
- HESELMANN, V., MAGER, A.-K., GOETZ, C., DETSCH, O., THEISGEN, H.-K., FRIESE, M., SCHWINDT, W., GOTTSCHALK, J. & KREMER, P. 2017. Accuracy of High-Field Intraoperative MRI in the Detectability of Residual Tumor in Glioma Grade IV Resections. *Treffsicherheit der Intraoperativen MR-Bildgebung (ioMRI) in der Nachweisbarkeit von Resttumorgewebe zur Resektion hochgradiger (Grad IV) Gliome.*, 189, 519-526.

- HILL, T. & LEWICKI, P. 2005. *Statistics: Methods and Applications*, Tulsa, OK, USA, Statsoft Inc.
- HINSHAW, W. S., BOTTOMLEY, P. A. & HOLLAND, G. N. 1977. Radiographic thin-section image of the human wrist by nuclear magnetic resonance. *Nature*, 270, 722-3.
- HUANG, Z., ZENG, H., HAMZAVI, I., MCLEAN, D. I. & LUI, H. 2001. Rapid near-infrared Raman spectroscopy system for real-time in vivo skin measurements. *Optics Letters*, 26, 1782-1784.
- JAKOLA, A. S., MYRMEL, K. S., KLOSTER, R., TORP, S. H., LINDAL, S., UNSGARD, G. & SOLHEIM, O. 2012. Comparison of a strategy favoring early surgical resection vs a strategy favoring watchful waiting in low-grade gliomas. *JAMA*, 308, 1881-8.
- JENKINSON, M. D., BARONE, D. G., BRYANT, A., VALE, L., BULBECK, H., LAWRIE, T. A., HART, M. G. & WATTS, C. 2018. Intraoperative imaging technology to maximise extent of resection for glioma. *Cochrane Database Syst Rev*, 1, CD012788.
- JERMYN, M., MOK, K., MERCIER, J., DESROCHES, J., PICHETTE, J., SAINT-ARNAUD, K., BERNSTEIN, L., GUIOT, M.-C., PETRECCA, K. & LEBLOND, F. 2015a. Intraoperative brain cancer detection with Raman spectroscopy in humans. 7, 1–10.
- JERMYN, M., MOK, K., MERCIER, J., DESROCHES, J., PICHETTE, J., SAINT-ARNAUD, K., BERNSTEIN, L., GUIOT, M.-C., PETRECCA, K. & LEBLOND, F. 2015b. Intraoperative brain cancer detection with Raman spectroscopy in humans. *Sci Transl Med*, 7, 274ra19.
- JI, M., LEWIS, S., CAMELO-PIRAGUA, S., RAMKISSOON, S. H., SNUDERL, M., VENNETI, S., FISHER-HUBBARD, A., GARRARD, M., FU, D., WANG, A. C., HETH, J. A., MAHER, C. O., SANAI, N., JOHNSON, T. D., FREUDIGER, C. W., SAGHER, O., XIE, X. S. & ORRINGER, D. A. 2015a. Detection of human brain tumor infiltration with quantitative stimulated Raman scattering microscopy. *Science translational medicine*, 7, 309ra163.
- JI, M., LEWIS, S., CAMELO-PIRAGUA, S., RAMKISSOON, S. H., SNUDERL, M., VENNETI, S., FISHER-HUBBARD, A., GARRARD, M., FU, D., WANG, A. C. & OTHERS 2015b. Detection of human brain tumor infiltration with quantitative stimulated Raman scattering microscopy. *Science translational medicine*, 7, 309ra163–309ra163.
- JI, M., ORRINGER, D. A., FREUDIGER, C. W., RAMKISSOON, S., LIU, X., LAU, D., GOLBY, A. J., NORTON, I., HAYASHI, M., AGAR, N. Y. R., YOUNG, G. S., SPINO, C., SANTAGATA, S., CAMELO-PIRAGUA, S., LIGON, K. L., SAGHER, O. & XIE, X. S. 2013. Rapid, Label-Free Detection of Brain Tumors with Stimulated Raman Scattering Microscopy. *Science Translational Medicine*, 5, 201ra119–201ra119.
- KALKANIS, S. N., KAST, R. E., ROSENBLUM, M. L., MIKKELSEN, T., YURGELEVIC, S. M., NELSON, K. M., RAGHUNATHAN, A., POISSON, L. M. & AUNER, G. W. 2014. Raman spectroscopy to distinguish grey matter, necrosis, and glioblastoma multiforme in frozen tissue sections. *Journal of Neuro-Oncology*, 116, 477–485.
- KAMEN, A., SUN, S., WAN, S., KLUCKNER, S., CHEN, T., GIGLER, A. M., SIMON, E., FLEISCHER, M., JAVED, M., DAALI, S., IGRESSA, A. & CHARALAMPAKI, P. 2016. Automatic Tissue Differentiation Based on Confocal Endomicroscopic Images for Intraoperative Guidance in Neurosurgery. *BioMed Research International*, 2016, 6183218.
- KANNAN, K., INAGAKI, A., SILBER, J., GOROVETS, D., ZHANG, J., KASTENHUBER, E. R., HEGUY, A., PETRINI, J. H., CHAN, T. A. & HUSE, J. T. 2012. Whole exome sequencing identifies ATRX mutation as a key molecular determinant in lower-grade glioma. *Oncotarget*, 3, 1194-1203.
- KHAMECHIAN, T., ALIZARGAR, J. & MAZOOCHI, T. 2012. The value of touch preparation for rapid diagnosis of brain tumors as an intraoperative consultation. *Iranian Journal of Medical Sciences*, 37, 105-111.
- KIERAN, M. W. 2014. Targeted treatment for sonic hedgehog-dependent medulloblastoma. *Neuro-Oncology*, 16, 1037-1047.
- KORIYAMA, S., NITTA, M., SHIOYAMA, T., KOMORI, T., MARUYAMA, T., KAWAMATA, T. & MURAGAKI, Y. 2018. Intraoperative Flow Cytometry Enables the Differentiation of Primary Central Nervous System Lymphoma from Glioblastoma. *World Neurosurgery*.
- KURIAN, K. M., JENKINSON, M. D., BRENNAN, P. M., GRANT, R., JEFFERIES, S., ROONEY, A. G., BULBECK, H., ERRIDGE, S. C., MILLS, S., MCBAIN, C., MCCABE, M. G., PRICE, S. J., MARINO, S., MOYES, E., QIAN, W., WALDMAN, A., VAQAS, B., KEATLEY, D., BURCHILL, P. & WATTS, C. Brain tumor research in the United Kingdom: current perspective and future challengesA strategy document from the NCRI Brain Tumor CSG. *Neuro-Oncology Practice*.
- LACROIX, M., ABI-SAID, D., FOURNEY, D. R., GOKASLAN, Z. L., SHI, W., DEMONTE, F., LANG, F. F., MCCUTCHEON, I. E., HASSENBUSCH, S. J., HOLLAND, E., HESS, K., MICHAEL, C., MILLER, D. & SAWAYA, R. 2001. A multivariate analysis of 416 patients with glioblastoma multiforme: prognosis, extent of resection, and survival. *Journal of Neurosurgery*, 95, 190-198.

- LAIKO, V. V., BALDWIN, M. A. & BURLINGAME, A. L. 2000. Atmospheric pressure matrix-assisted laser desorption/ionization mass spectrometry. *Analytical Chemistry*, 72, 652-657.
- LAU, D., HERVEY-JUMPER, S. L., CHANG, S., MOLINARO, A. M., MCDERMOTT, M. W., PHILLIPS, J. J. & BERGER, M. S. 2016. A prospective Phase II clinical trial of 5-aminolevulinic acid to assess the correlation of intraoperative fluorescence intensity and degree of histologic cellularity during resection of high-grade gliomas. *Journal of neurosurgery*, 124, 1300-9.
- LEE, J.-K., WANG, J., SA, J. K., LADEWIG, E., LEE, H.-O., LEE, I.-H., KANG, H. J., ROSENBLOOM, D. S., CAMARA, P. G., LIU, Z., VAN NIEUWENHUIZEN, P., JUNG, S. W., CHOI, S. W., KIM, J., CHEN, A., KIM, K.-T., SHIN, S., SEO, Y. J., OH, J.-M., SHIN, Y. J., PARK, C.-K., KONG, D.-S., SEOL, H. J., BLUMBERG, A., LEE, J.-I., IAVARONE, A., PARK, W.-Y., RABADAN, R. & NAM, D.-H. 2017. Spatiotemporal genomic architecture informs precision oncology in glioblastoma. *Nature genetics*, 49, 594-599.
- LESLIE, D. G., KAST, R. E., POULIK, J. M., RABAH, R., SOOD, S., AUNER, G. W. & KLEIN, M. D. 2012. Identification of pediatric brain neoplasms using raman spectroscopy. *Pediatric Neurosurgery*, 48, 109-117.
- LEWIS, P. W., ELSAESSER, S. J., NOH, K.-M., STADLER, S. C. & ALLIS, C. D. 2010. Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. *Proceedings of the National Academy of Sciences*, 107, 14075-14080.
- LIAU, L. M., ASHKAN, K., TRAN, D. D., CAMPIAN, J. L., TRUSHEIM, J. E., COBBS, C. S., HETH, J. A., SALACZ, M., TAYLOR, S., D'ANDRE, S. D., IWAMOTO, F. M., DROPCHO, E. J., MOSHEL, Y. A., WALTER, K. A., PILLAINAYAGAM, C. P., AIKEN, R., CHAUDHARY, R., GOLDLUST, S. A., BOTA, D. A., DUIC, P., GREWAL, J., ELINZANO, H., TOMS, S. A., LILLEHEI, K. O., MIKKELSEN, T., WALPERT, T., ABRAM, S. R., BRENNER, A. J., BREM, S., EWEND, M. G., KHAGI, S., PORTNOW, J., KIM, L. J., LOUDON, W. G., THOMPSON, R. C., AVIGAN, D. E., FINK, K. L., GEOFFROY, F. J., LINDHORST, S., LUTZKY, J., SLOAN, A. E., SCHACKERT, G., KREX, D., MEISEL, H. J., WU, J., DAVIS, R. P., DUMA, C., ETAME, A. B., MATHIEU, D., KESARI, S., PICCIONI, D., WESTPHAL, M., BASKIN, D. S., NEW, P. Z., LACROIX, M., MAY, S. A., PLUARD, T. J., TSE, V., GREEN, R. M., VILLANO, J. L., PEARLMAN, M., PETRECCA, K., SCHULDER, M., TAYLOR, L. P., MAIDA, A. E., PRINS, R. M., CLOUGHESY, T. F., MULHOLLAND, P. & BOSCH, M. L. 2018. First results on survival from a large Phase 3 clinical trial of an autologous dendritic cell vaccine in newly diagnosed glioblastoma. *J Transl Med*, 16, 142.
- LOUIS, D., OHGAKI, H., WIESTLER, O. & CAVENEE, W. 2016. *WHO Classification of Tumours of the Central Nervous System*.
- MA, R., CHARI, A., BRENNAN, P., COLLABORATION, B. N. T. R., WATTS, C., NEUROCHIRURGIE, D. G. F. & SURGEONS, S. O. B. N. 2017. Residual enhancing disease after surgery for glioblastoma multiforme: National service evaluation of practice in the United Kingdom. German Medical Science GMS Publishing House.
- MANDONNET, E. & DUFFAU, H. 2017. The Concept of Onco-Functional Balance in the Management of DLGG. *Diffuse Low-Grade Gliomas in Adults*. Springer, Cham.
- MANDONNET, E., WAGER, M., ALMAIRAC, F., BARON, M.-H., BLONSKI, M., FREYSCHLAG, C. F., BARONE, F., FONTAINE, D., PALLUD, J., HEGI, M., VIEGAS, C., ZETTERLING, M., SPENA, G., GOODDEN, J., RUTTEN, G.-J., TAILLANDIER, L., FOROGLU, N., DARLIX, A., SKRAP, M., MARTINO, J., VON CAMPE, G., MADADAKI, C., GAYAT, E., DE WITT HAMER, P., GIL ROBLES, S., SARUBBO, S., SANTORIUS, T., BELLO, L., FORSTER, M.-T. & DUFFAU, H. 2017. Survey on current practice within the European Low-Grade Glioma Network: where do we stand and what is the next step? *Neuro-Oncology Practice*, 4, 241-247.
- MARKO, N. F., WEIL, R. J., SCHROEDER, J. L., LANG, F. F., SUKI, D. & SAWAYA, R. E. 2014. Extent of resection of glioblastoma revisited: personalized survival modeling facilitates more accurate survival prediction and supports a maximum-safe-resection approach to surgery. *J Clin Oncol*, 32, 774-82.
- MCCULLOCH, P., ALTMAN, D. G., CAMPBELL, W. B., FLUM, D. R., GLASZIOU, P., MARSHALL, J. C. & NICHOLL, J. 2009. No surgical innovation without evaluation: the IDEAL recommendations. *The Lancet*, 374, 1105-1112.
- MCGREGOR, H. C., SHORT, M. A., MCWILLIAMS, A., SHAIPIANICH, T., IONESCU, D. N., ZHAO, J., WANG, W., CHEN, G., LAM, S. & ZENG, H. 2017. Real-time endoscopic Raman spectroscopy for in vivo early lung cancer detection. *Journal of Biophotonics*, 10, 98-110.
- MCLENDON, R., FRIEDMAN, A., BIGNER, D., MEIR, E. G. V., BRAT, D. J., MASTROGIANAKIS, G. M., OLSON, J. J., MIKKELSEN, T., LEHMAN, N., ALDAPE, K., YUNG, W. K. A., BOGLER, O., VANDENBERG, S., BERGER, M., PRADOS, M., MUZNY, D., MORGAN, M., SCHERER, S., SABO, A., NAZARETH, L., LEWIS, L., HALL, O., ZHU, Y., REN, Y., ALVI, O., YAO, J., HAWES, A., JHANGIANI, S., FOWLER, G., LUCAS, A. S., KOVAR, C., CREE, A., DINH, H., SANTIBANEZ, J., JOSHI, V., GONZALEZ-GARAY, M. L., MILLER, C. A., MILOSAVLJEVIC, A., DONEHOWER, L., WHEELER, D. A., GIBBS, R. A., CIBULSKIS, K., SOUGNEZ, C., FENNELL, T., MAHAN,

- S., WILKINSON, J., ZIAUGRA, L., ONOFRIO, R., BLOOM, T., NICOL, R., ARDLIE, K., BALDWIN, J., GABRIEL, S., LANDER, E. S., DING, L., FULTON, R. S., MCLELLAN, M. D., WALLIS, J., LARSON, D. E., SHI, X., ABBOTT, R., FULTON, L., CHEN, K., KOBOLDT, D. C., WENDL, M. C., MEYER, R., TANG, Y., LIN, L., OSBORNE, J. R., DUNFORD-SHORE, B. H., MINER, T. L., DELEHAUNTY, K., MARKOVIC, C., SWIFT, G., COURTNEY, W., POHL, C., ABBOTT, S., HAWKINS, A., LEONG, S., HAIPEK, C., SCHMIDT, H., WIECHERT, M., VICKERY, T., SCOTT, S., DOOLING, D. J., CHINWALLA, A., WEINSTOCK, G. M., MARDIS, E. R., WILSON, R. K., GETZ, G., WINCKLER, W., VERHAAK, R. G. W., LAWRENCE, M. S., O'KELLY, M., ROBINSON, J., ALEXE, G., BEROUKHIM, R., CARTER, S., CHIANG, D., GOULD, J., et al. 2008. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*, 455, 1061-1068.
- MIRNEZAMI, R., NICHOLSON, J. & DARZI, A. 2012. Preparing for Precision Medicine. *New England Journal of Medicine*, 366, 489-491.
- NORTHCOTT, P. A., KORSHUNOV, A., WITT, H., HIELSCHER, T., EBERHART, C. G., MACK, S., BOUFFET, E., CLIFFORD, S. C., HAWKINS, C. E., FRENCH, P., RUTKA, J. T., PFISTER, S. & TAYLOR, M. D. 2011. Medulloblastoma Comprises Four Distinct Molecular Variants. *Journal of Clinical Oncology*, 29, 1408-1414.
- ORRINGER, D. A., PANDIAN, B., NIKNAFS, Y. S., HOLLON, T. C., BOYLE, J., LEWIS, S., GARRARD, M., HERVEY-JUMPER, S. L., GARTON, H. J. L., MAHER, C. O., HETH, J. A., SAGHER, O., WILKINSON, D. A., SNUDERL, M., VENNETI, S., RAMKISSOON, S. H., MCFADDEN, K. A., FISHER-HUBBARD, A., LIEBERMAN, A. P., JOHNSON, T. D., XIE, X. S., TRAUTMAN, J. K., FREUDIGER, C. W. & CAMELO-PIRAGUA, S. 2017. Rapid intraoperative histology of unprocessed surgical specimens via fibre-laser-based stimulated Raman scattering microscopy. *Nature Biomedical Engineering*, 1, 0027.
- OSTROM, Q. T., GITTLEMAN, H., BLANK, D., M, P., FINLAY, J. L., GURNEY, J. G., MCKEAN-COWDIN, R., STEARNS, D. S., WOLFF, J. E., LIU, M., WOLINSKY, Y., KRUCHKO, C. & BARNHOLTZ-SLOAN, J. S. 2016. American Brain Tumor Association Adolescent and Young Adult Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2008-2012. *Neuro-Oncology*, 18, i1-i50.
- OSTROM, Q. T., GITTLEMAN, H., FULOP, J., LIU, M., BLANDA, R., KROMER, C., WOLINSKY, Y., KRUCHKO, C. & BARNHOLTZ-SLOAN, J. S. 2015. CBTRUS Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2008-2012. *Neuro-Oncology*, 17, iv1-iv62.
- PENFIELD, W. & BOLDREY, E. 1937. SOMATIC MOTOR AND SENSORY REPRESENTATION IN THE CEREBRAL CORTEX OF MAN AS STUDIED BY ELECTRICAL STIMULATION. *Brain*, 60, 389-443.
- PHILLIPS, N. 2018. *Cranial Neurosurgery, GIRFT Programme National Speciality Report*, Royal National Orthopaedic Hospital and NHS Improvement.
- REES, C. J., RAJASEKHAR, P. T., WILSON, A., CLOSE, H., RUTTER, M. D., SAUNDERS, B. P., EAST, J. E., MAIER, R., MOORGHEN, M., MUHAMMAD, U., HANCOCK, H., JAYAPRAKASH, A., MACDONALD, C., RAMADAS, A., DHAR, A. & MASON, J. M. 2017. Narrow band imaging optical diagnosis of small colorectal polyps in routine clinical practice: the Detect Inspect Characterise Resect and Discard 2 (DISCARD 2) study. *Gut*, 66, 887-895.
- RITSCHER, K., PECHLIVANIS, I. & WINTER, S. 2015. Brain tumor classification on intraoperative contrast-enhanced ultrasound. *International Journal of Computer Assisted Radiology and Surgery*, 10, 531-540.
- SAHOOLIZADEH, A. H., HEIDARI, B. Z. & DEGHANI, C. H. 2008. A new face recognition method using PCA, LDA and neural network. *International Journal of Computer Science and Engineering*, 2, 218-223.
- SANAI, N. & BERGER, M. S. 2008. GLIOMA EXTENT OF RESECTION AND ITS IMPACT ON PATIENT OUTCOME. *Neurosurgery*, 62, 753-766.
- SANAI, N., POLLEY, M.-Y., MCDERMOTT, M. W., PARSA, A. T. & BERGER, M. S. 2011. An extent of resection threshold for newly diagnosed glioblastomas. *Journal of Neurosurgery*, 115, 3-8.
- SANTAGATA, S., EBERLIN, L. S., NORTON, I., CALLIGARIS, D., FELDMAN, D. R., IDE, J. L., LIU, X., WILEY, J. S., VESTAL, M. L., RAMKISSOON, S. H., ORRINGER, D. A., GILL, K. K., DUNN, I. F., DIAS-SANTAGATA, D., LIGON, K. L., JOLESZ, F. A., GOLBY, A. J., COOKS, R. G. & AGAR, N. Y. R. 2014. Intraoperative mass spectrometry mapping of an onco-metabolite to guide brain tumor surgery. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 11121-11126.
- SAVITZKY, A. & GOLAY, M. J. E. 1964. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, 36, 1627-1639.
- SCHÄFER, K.-C., DÉNES, J., ALBRECHT, K., SZANISZLÓ, T., BALOG, J., SKOUMAL, R., KATONA, M., TÓTH, M., BALOGH, L. & TAKÁTS, Z. 2009. In Vivo, In Situ Tissue Analysis Using Rapid Evaporative Ionization Mass Spectrometry. *Angewandte Chemie International Edition*, 48, 8240-8242.
- SENDERS, J. T., ZAKI, M. M., KARHADE, A. V., CHANG, B., GORMLEY, W. B., BROEKMAN, M. L., SMITH, T. R. &

- ARNAOUT, O. 2018. An introduction and overview of machine learning in neurosurgical care. *Acta Neurochirurgica*, 160, 29-38.
- SENG, P., ROLAIN, J.-M., FOURNIER, P. E., LA SCOLA, B., DRANCOURT, M. & RAOULT, D. 2010. MALDI-TOF-mass spectrometry applications in clinical microbiology. *Future microbiology*, 5, 1733-1754.
- SHIOYAMA, T., MURAGAKI, Y., MARUYAMA, T., KOMORI, T. & ISEKI, H. 2013. Intraoperative flow cytometry analysis of glioma tissue for rapid determination of tumor presence and its histopathological grade. *Journal of Neurosurgery*, 118, 1232-1238.
- SHORT, M. A., LAM, S., MCWILLIAMS, A., ZHAO, J., LUI, H. & ZENG, H. 2008. Development and preliminary results of an endoscopic Raman probe for potential in vivo diagnosis of lung cancers. *Optics Letters*, 33, 711-713.
- SHORT, M. A., LUI, H., MCLEAN, D., ZENG, H., ALAJLAN, A. & CHEN, X. K. 2006. Changes in nuclei and peritumoral collagen within nodular basal cell carcinomas via confocal micro-Raman spectroscopy. *J Biomed Opt*, 11, 34004.
- SHORT, M. A., WANG, W., TAI, I. T. & ZENG, H. 2016. Development and in vivo testing of a high frequency endoscopic Raman spectroscopy system for potential applications in the detection of early colonic neoplasia. *Journal of Biophotonics*, 9, 44-48.
- SNEDDEN, W. & PARKER, R. B. 1971. Determination of volatile constituents of human blood and tissue specimens by quantitative high-resolution mass spectrometry. *Analytical Chemistry*, 43, 1651-1656.
- SOTTORIVA, A., KANG, H., MA, Z., GRAHAM, T. A., SALOMON, M. P., ZHAO, J., MARJORAM, P., SIEGMUND, K., PRESS, M. F., SHIBATA, D. & CURTIS, C. 2015. A Big Bang model of human colorectal tumor growth. *Nat Genet*, 47, 209-16.
- SOTTORIVA, A., SPITERI, I., PICCIRILLO, S. G. M., TOULLOUMIS, A., COLLINS, V. P., MARIONI, J. C., CURTIS, C., WATTS, C. & TAVARÉ, S. 2013. Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proceedings of the National Academy of Sciences*, 110, 4009-4014.
- ST JOHN, E. R., AL-KHUDAIRI, R., ASHRAFIAN, H., ATHANASIOU, T., TAKATS, Z., HADJIMINAS, D. J., DARZI, A. & LEFF, D. R. 2017a. Diagnostic Accuracy of Intraoperative Techniques for Margin Assessment in Breast Cancer Surgery: A Meta-analysis. *Annals of Surgery*, 265, 300.
- ST JOHN, E. R., BALOG, J., MCKENZIE, J. S., ROSSI, M., COVINGTON, A., MUIRHEAD, L., BODAI, Z., ROSINI, F., SPELLER, A. V. M., SHOUSA, S., RAMAKRISHNAN, R., DARZI, A., TAKATS, Z. & LEFF, D. R. 2017b. Rapid evaporative ionisation mass spectrometry of electrosurgical vapours for the identification of breast pathology: towards an intelligent knife for breast cancer surgery. *Breast Cancer Research : BCR*, 19.
- STRITTMATTER, N., JONES, E. A., VESELKOV, K. A., REBEC, M., BUNDY, J. G. & TAKATS, Z. 2013. Analysis of Intact Bacteria using Rapid Evaporative Ionisation Mass Spectrometry. *Chem. Commun.*, 49, 6188-6190.
- STRITTMATTER, N., REBEC, M., JONES, E. A., GOLF, O., ABDOLRASOULI, A., BALOG, J., BEHREND, V., VESELKOV, K. A. & TAKATS, Z. 2014. Characterization and identification of clinically relevant microorganisms using rapid evaporative ionization mass spectrometry. *Anal. Chem.*, 86, 6555-6562.
- STUMMER, W. 2015. Fluorescein in brain metastasis and glioma surgery. *Acta Neurochirurgica*, 157, 2199-2200.
- STUMMER, W., PICHLMEIER, U., MEINEL, T., WIESTLER, O. D., ZANELLA, F., REULEN, H.-J., GROUP, A.-G. S. & OTHERS 2006. Fluorescence-guided surgery with 5-aminolevulinic acid for resection of malignant glioma: a randomised controlled multicentre phase III trial. *The lancet oncology*, 7, 392-401.
- STUMMER, W., REULEN, H. J., NOVOTNY, A., STEPP, H. & TONN, J. C. 2003. Fluorescence-guided resections of malignant gliomas--an overview. *Acta neurochirurgica. Supplement*, 88, 9-12.
- TANINO, M., SASAJIMA, T., NANJO, H., AKESAKA, S., KAGAYA, M., KIMURA, T., ISHIDA, Y., ODA, M., TAKAHASHI, M., SUGAWARA, T., YOSHIOKA, T., NISHIHARA, H., AKAGAMI, Y., GOTO, A., MINAMIYA, Y., TANAKA, S. & GROUP, R.-I. S. 2015. Rapid immunohistochemistry based on alternating current electric field for intraoperative diagnosis of brain tumors. *Brain tumor pathology*, 32, 12-9.
- THON, N., KRETH, S. & KRETH, F.-W. 2013. Personalized treatment strategies in glioblastoma: MGMT promoter methylation status. *Oncotargets and therapy*, 6, 1363-1372.
- VERHAAK, R. G. W., HOADLEY, K. A., PURDOM, E., WANG, V., QI, Y., WILKERSON, M. D., MILLER, C. R., DING, L., GOLUB, T., MESIROV, J. P., ALEXE, G., LAWRENCE, M., O'KELLY, M., TAMAYO, P., WEIR, B. A., GABRIE, S., WINCKLER, W., GUPTA, S., JAKKULA, L., FEILER, H. S., HODGSON, J. G., JAMES, C. D., SARKARIA, J. N., BRENNAN, C., KAHN, A., SPELLMAN, P. T., WILSON, R. K., SPEED, T. P., GRAY, J. W., MEYERSON, M., GETZ, G., PEROU, C. M. & HAYES, D. N. 2010. An integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR and NF1. *Cancer cell*, 17, 98.
- VOGELBAUM, M. A. & AGHI, M. K. 2015. Convection-enhanced delivery for the treatment of glioblastoma.

Neuro-Oncology, 17, ii3-ii8.

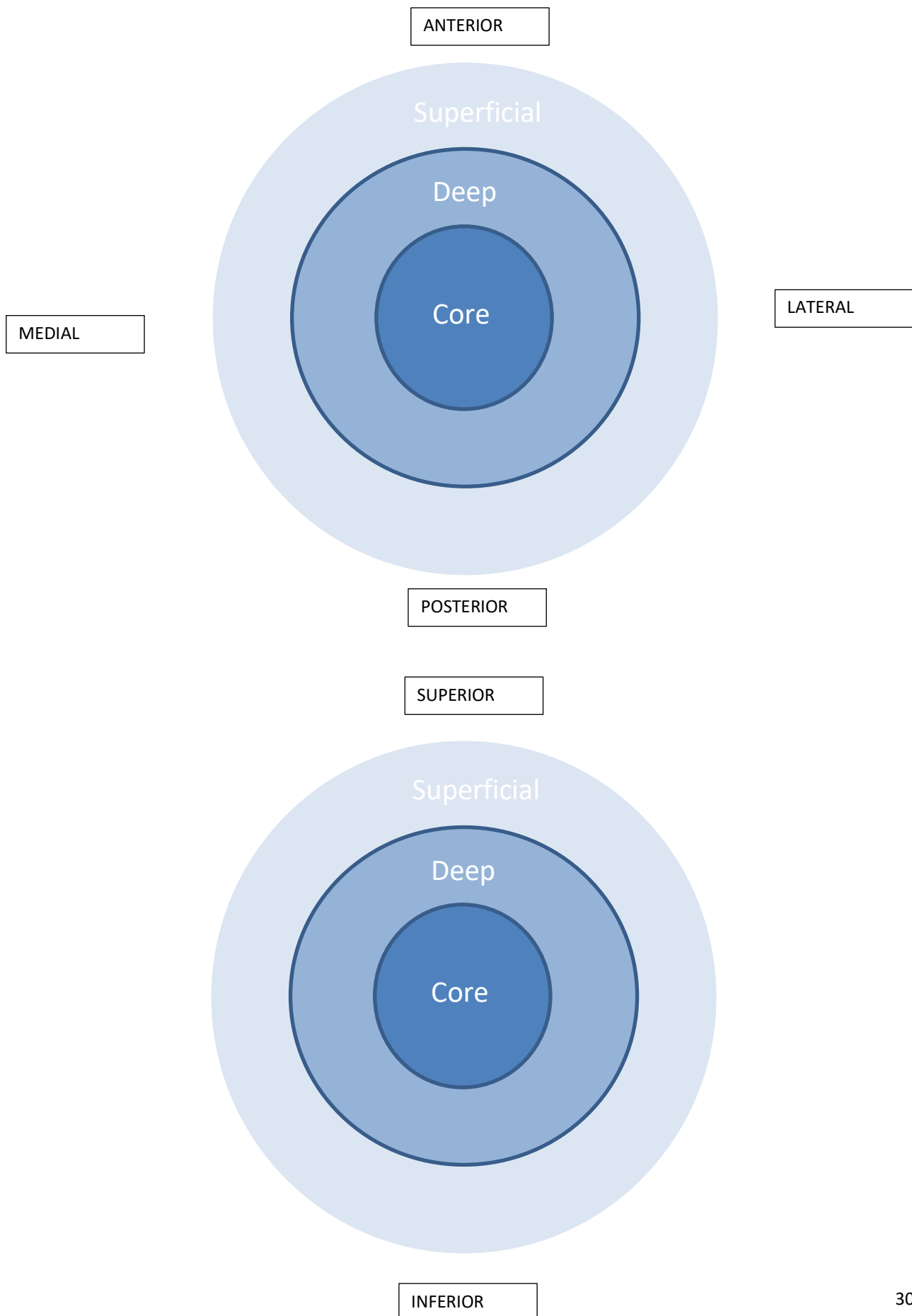
- WATTS, C. 2016. Targeting residual enhancing disease in glioma in UK practice: a survey of interest and resourcing for surgical oncology trials. *British Journal of Neurosurgery*, 30, 478.
- WHITING, P. F., RUTJES, A. W., WESTWOOD, M. E., MALLETT, S., DEEKS, J. J., REITSMA, J. B., LEEFLANG, M. M., STERNE, J. A., BOSSUYT, P. M. & GROUP, Q.-. 2011. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*, 155, 529-36.
- WIDHALM, G., KIESEL, B., WOEHRER, A., TRAUB-WEIDINGER, T., PREUSSER, M., MAROSI, C., PRAYER, D., HAINFELLNER, J. A., KNOSP, E. & WOLFSBERGER, S. 2013. 5-Aminolevulinic acid induced fluorescence is a powerful intraoperative marker for precise histopathological grading of gliomas with non-significant contrast-enhancement. *PLoS one*, 8, e76988.
- YANG, D., RAO, G., MARTINEZ, J., VEERARAGHAVAN, A. & RAO, A. 2015. Evaluation of tumor-derived MRI-texture features for discrimination of molecular subtypes and prediction of 12-month survival status in glioblastoma. *Med Phys*, 42, 6725-35.
- ZANELLO, M., POULON, F., PALLUD, J., VARLET, P., HAMZE, H., ABI LAHOUD, G., ANDREIUOLO, F., IBRAHIM, A., PAGES, M., CHRETIEN, F., DI ROCCO, F., DEZAMIS, E., NATAF, F., TURAK, B., DEVAUX, B. & ABI HAIDAR, D. 2017. Multimodal optical analysis discriminates freshly extracted human sample of gliomas, metastases and meningiomas from their appropriate controls. *Sci Rep*, 7, 41724.
- ZHAO, J., LUI, H., MCLEAN, D. I. & ZENG, H. 2007. Automated Autofluorescence Background Subtraction Algorithm for Biomedical Raman Spectroscopy. *Applied Spectroscopy*, 61, 1225-1232.
- ZHAO, J., LUI, H., MCLEAN, D. I. & ZENG, H. 2016. Automated Autofluorescence Background Subtraction Algorithm for Biomedical Raman Spectroscopy. *Applied Spectroscopy*.
- ZHAO, J., SHORT, M., BRAUN, T., LUI, H., MCLEAN, D. & ZENG, H. 2014. Clinical Raman measurements under special ambient lighting illumination. *Journal of Biomedical Optics*, 19, 111609-111609.

7. Appendix

7.1. Appendix 1: Raman In vivo Data Entry Sheet

[illegible]

7.2. Appendix 2: in vivo Specimen Labelling Guide



7.3. Appendix 3: iKnife Setup, Calibration and Troubleshooting

The instrument was set up in a location determined as previously described. Before switching on the instrument, the REIMS machine was connected to mains power (it is not advisable to plug the instrument into a multiplug strip) and connected to a 7 bar gas supply, connected to the operating theatre gas supply.

- The connections were then checked by opening the flow on the cylinder and checking for leaks in connectors. If any were found, airflow was stopped and connections were re-checked. This process was repeated until a safe, closed circuit was established.

Once connected safely to the pneumatised air system in theatres as a closed loop the system can then be switched on.

- Plug in the instrument: automatic startup.
- Switch on electronics by lifting the electronics switch at the back or side of the instrument to the upright position. Pull switch out slightly and raise. An example is shown below.
- Allow 5 minutes for the embedded PC to initialise and then start MassLynx software. This is an important step. Failing to wait the allotted time could result in the computer and instrument not communicating.
- From the instrument tab (1) choose MS Tune (2)
- Open the MS Tune window that has appeared in the toolbar
- Go to vacuum pump
- Go to 'view vacuum' and wait for the pressure displays to reach green and turbo speed fore-line pumps to reach 100%. Allow the instrument to pump down until the TOF pressure is at least 1.2×10^{-6} mbar and the backing pressure is between 1.3 and 2 mbar before attempting to switch into operate mode (this may take up to four hours, even longer if the instrument has been off some time).
- If the instrument has been switched off completely, go to set up system view and enter password provided by instrument manufacturer ('access'). Then choose set up voltages on. You will now be able to operate the mass spectrometer.

Preparing instrument for data acquisition:

- Switch on gas supply from main valve.
- Put the instrument into operational state. The coloured box will turn from red to green.

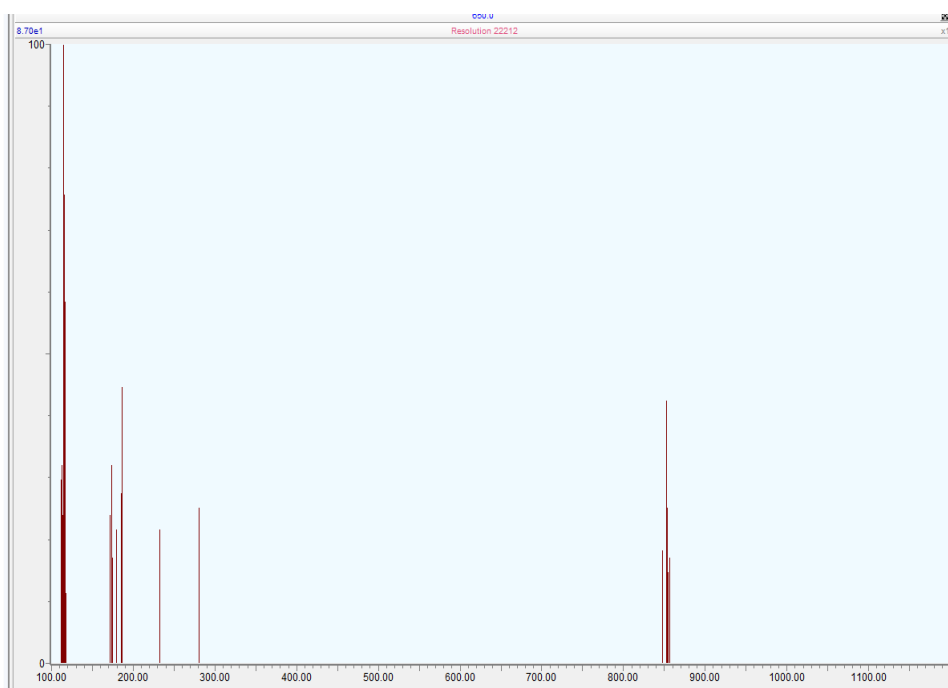


Figure 168. Background signal on REIMS machine.

- Background signal should be seen as shown in Figure 168. If not, go to system reinitialise.
- Select the mode of operation: negative ionization mode [1], sensitivity resolution [2], MS mode [3] and trap gas [4]. In scan conditions select 1.0 scan time [5].

Switching off the instrument

- It should be noted that a mass spectrometer should always be on and pumping vacuum. The instrument should be vented and/or turned off only when necessary.
- On the embedded instrument computer, go to MassLynx software → on Instrument tab (1) select MS tune (2) → Vacuum → Vent. Wait for approximately 2 minutes for the three turbo speeds to drop to 2%, then close the program.
- On the main computer home screen, go to start and restart. It is advisable to not select 'shut down' as this will turn off the computer inside the instrument and to turn this back on will require opening up a side panel. When the computer has gone to the black 'analogue' screen, switch off the mains power.
- Switch off the instrument electronics (on the left hand side panel - bottom switch - down for off).

Preparing instrument for *in vivo* collection involves preparing the IPA solution as follows:

- Decant some IPA into a clean beaker. Using a pipette and measuring cylinder, transfer 40.5ml of the decanted IPA into a 100ml volumetric flask.
- Into a second clean beaker, decant roughly 10ml of LCMS grade water. Using a 1000µl pipette, transfer 4.5ml of LCMS grade water into the same volumetric flask, cap and invert to mix. This produces a 90:10 IPA:Water solution.
- Using a 200µl pipette, transfer 56µl of 4M Sodium formate solution into the 90:10 IPA:Water solution, cap and invert to mix.

- Label the solution 5mM Sodium formate, plus the date of preparation, expiry date and initials. Store for up to 4 weeks at between 2-8°C.

Leucine Enkephalin Lockmass Solution of 1ng/μL was prepared as follows:

- Transfer the full 25 mg of Leucine Enkephalin powder into a 50mL volumetric flask by washing it out with IPA. After ensuring all the powder has been transferred, fill the flask to volume with IPA. This results in a stock solution with a concentration of 500 ng/μL. This should be stored in a -20°C freezer.
- Wrap the flask in metal foil to avoid light exposure and label the flask as 500 ng/μL Leu Enk Stock solution, plus the date of preparation and your initials (alternatively, use amber glassware).
- Using an air displacement pipette, transfer 200 μL of the stock solution into a 100 mL volumetric flask and fill to volume with IPA. This results in a solution with a concentration of 1 ng/μL.
- Label flask as '1 ng/μL Leu Enk Solution', along with the date of preparation and your initials.
- Other concentrations of Leu Enk can be prepared from the stock solution using the calculation:

$$C1 \times V1 = C2 \times V2$$
- For example, if the required concentration is 5 ng/μL (C2) in 50 mL (V2) then the volume required (V1) is 0.5 mL or 500 μL.

Tissue was stored using the following protocol:

- On receipt of a sample for REIMS analysis, it must be assigned a unique sample number appropriate to the project it is for, e.g. for breast tissue this is an IKB number. This should not be the same as a patient's hospital number.
- If the sample is received fresh from the pathologist, it must be immediately refrigerated and can be stored here for up to 4 hours before analysis. After 4 hours it is not recommended to analyse as fresh tissue and must be transferred to a -80°C freezer.
- If the tissue is received frozen, after labelling it must be stored in a -80°C freezer.
- For tissue database purposes, the sample must be logged onto the appropriate project database with unique reference number and storage location.
- If analysing fresh tissue, remove from refrigerator and remove a portion for analysis by REIMS. Transfer the remaining tissue to the -80°C freezer and log onto the appropriate project database.
- If analysing frozen tissue, remove from the freezer and remove a portion for analysis by REIMS. Return the remaining tissue to the -80°C freezer in the appropriate location and record that tissue has been removed for analysis if necessary.
- Allow tissue to reach room temperature before analysis and if necessary, keep moist by wetting with deionised water.
- Label a histo-cassette with the sample's unique identifier.
- After analysis, place the analysed tissue in the labelled histo-cassette and store in 10% formalin solution for subsequent pathological tissue classification.

7.3.1. Instrument calibration

- Fill a 5ml Hamilton syringe with 5mM NaF solution
- o Ensure the syringe is wet with a small amount of the solution before assembling
- Attach syringe to PEEK tubing and push a small amount of solution through the tubing.
- o Ensure there are no air bubbles in the syringe by holding the syringe upright and placing a piece of blue roll at the end of the tubing to catch any solution while pushing the syringe slowly.
- Secure the syringe to the syringe pump and move the baseplate in position, i.e. touching the back of the syringe.
- Set the syringe pump to the correct syringe size and flow rate – in this case the flow rate should be 0.20ml/min. This can be changed if the signal is a bit weak.
- Ensure gas supply is switched on, take instrument out of standby by clicking where shown:
 - The instrument will automatically turn on the API gas, turn it off.
 - Turn on the coil
- Place the calibration capillary (the longer of the two capillaries) into the inlet on the Venturi interface.
- Connect the free end of the PEEK tubing to the capillary using a finger-tight connector. Do not screw this in too tight as it restricts the flow of the solution. If the flow is restricted you will not see any signal, or very limited signal. Loosen the finger tight connector a small amount and the signal should improve.
- Press start on the syringe pump.
- Wait until you can see a fair amount of signal in the MS Tune page, shown below, and press the play button.
- Filename should be in the format of YYYYMMDD_CAL_NEG_SENS
- Let the file run for approximately two minutes (this can be timed using the chromatogram window. To observe the run in 'realtime' click the clock icon on the toolbar) and press the stop button. When the software asks 'Really stop acquisition' click 'ok'.
- Stop the syringe pump.
- From the chromatogram window, right click on the mouse and drag across the chromatogram.
- This will open a spectrum window:
- Click process, automatic peak detection. A window will open and the following options should be ticked, click ok:
 - This will open a new spectrum. Click file, save spectrum, ok.
- Return back to the MS Tune page. Click Setup Calibration Calibrate TOF. The calibration window will open.
- Ensure that the dropdown box says ESI_NaFormate_Neg and click Calibrate Create Calibration. A window will open. Select the file that has just been run.
- Click history and select the second AccMass option. Click ok.
- The following window shown in Figure 169 will open up.

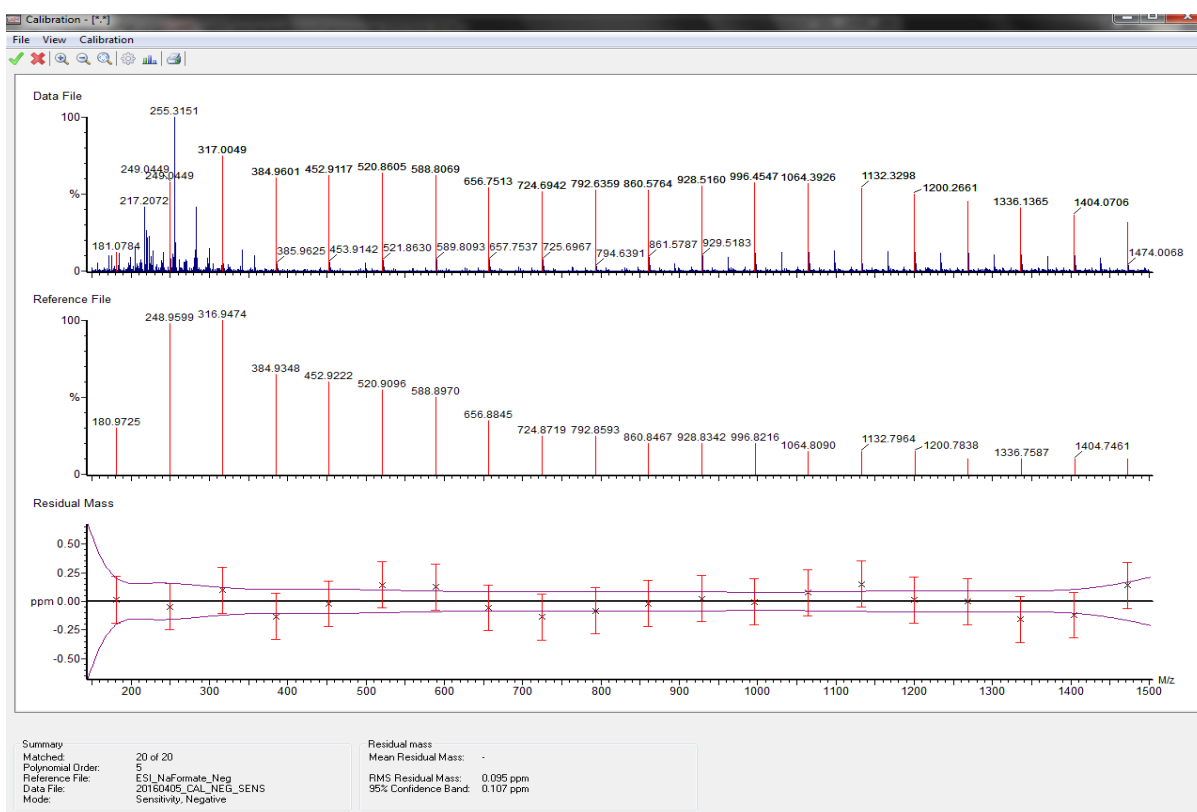


Figure 169. Calibration screen on the iKnife

- RMS residual at 95% confidence should be below 3ppm.
- o Check that the masses, especially in the 600 – 1000 m/z region, are matched and found. If a mass is greyed out, then it wasn't found.
- o If a specific mass is deemed to be an outlier (i.e. the RSM residual is higher than 3ppm), it can be removed. To do this, right click the mass in the ref file and double right click on the matching mass above it.
- o If the mass removed does not improve the RMS residual, it can be reinstated. To reinstate a mass that was removed, click view, reselect peaks.
- Once the calibration is satisfactory, click the green tick.
- Return to the calibration window and click save as YYYYMMDD_neg_sens and save.
- A window will open and ask which parameters the new settings apply to. Select both tuning parameters and system settings, click ok.
- To check that this calibration is being used by the instrument, the calibration file name should be visible at the bottom right of the MS Tune page.

To check calibration the following steps should be performed:

- To check the calibration press start on the syringe pump.
- Wait until you can see a fair amount of signal in the MS Tune page and press the play button.
- Filename should be in the format of YYYYMMDD_CAL_NEG_SENS_CHK
- Let the file run for approximately one minute and press the stop button. When the software asks 'Really stop acquisition' click 'ok'.
- From the chromatogram window, right click on the mouse and drag across the chromatogram. This will open a spectrum window.
- Click process, automatic peak detection. A window will open and the lockmass

of 724.8718 entered.

- Click Edit Copy spectrum list.
- Open up MassDiff from the start menu
- Click Open Ref file. This should be ESI_NaFormate_Neg
- Click the clipboard icon.
- The overall status should be pass, as shown in green in Figure 170.

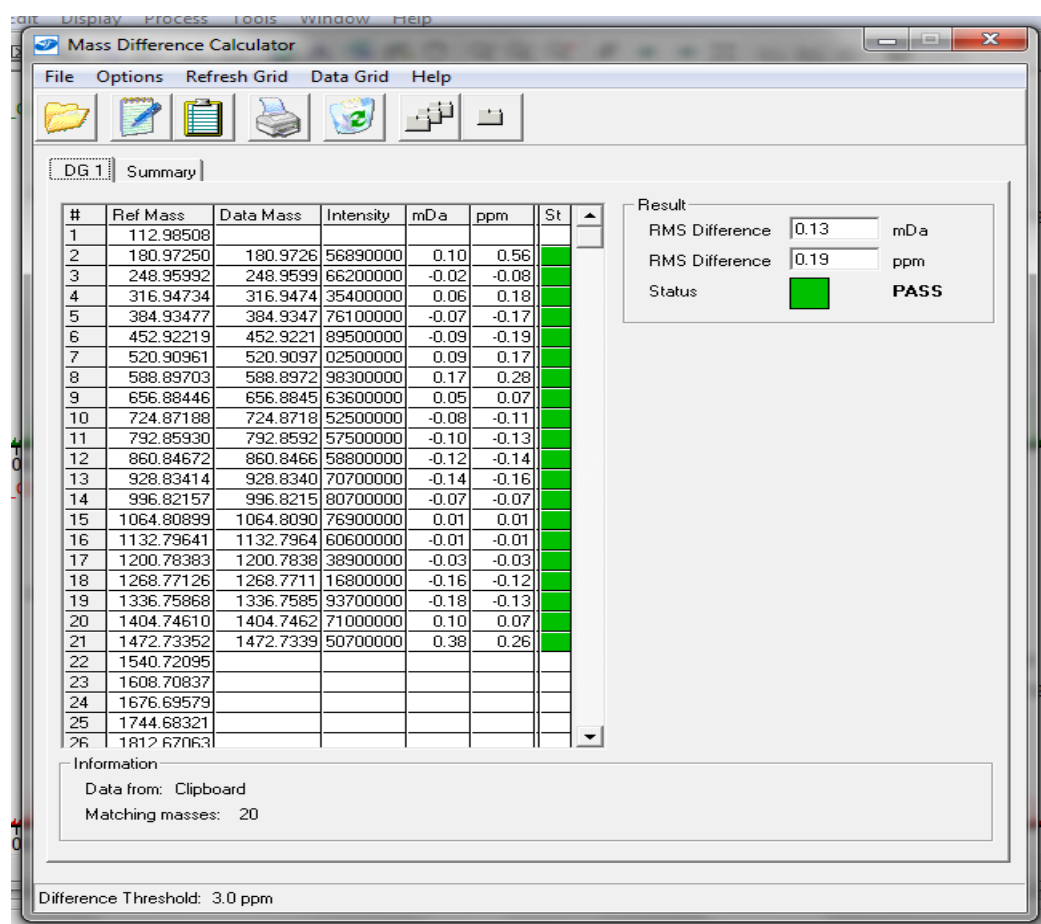


Figure 170. Calibration check pass screen for the iKnife.

If the calibration check fails, another check file will need to be run and checked.

Note: Before running samples, whether in vivo or ex vivo, it is recommended that IPA be run through the system to clear out any remaining NaF solution – this can be run at the same flow rate and no file needs to be collected. It is advisable to have two separate PEEK tubing sections; one exclusively for calibration solution and one for solvent (i.e. IPA, Leu Enk and the like) as cross contamination could occur.

Instrument maintenance

The procedure below details how to clean and maintain the instrument.

Cleaning the Interface and Stepwave

- Before removing the interface and StepWave, the instrument must be vented. On the instrument computer, go to MassLynx software → on Instrument tab, select MS tune → Vacuum → Vent. Wait for approximately 3 minutes for the three turbo

speeds to drop to 2%.

- Remove the Venturi chamber from the interface by unscrewing the black bolt and gently easing away from the interface housing. Put any IPA capillary inlet pieces to one side for cleaning. Remove the Venturi inlet piece and disassemble by unscrewing the two component parts. Put all pieces except for the O-ring into a small beaker for cleaning.
- Unscrew the nut that holds the IPA capillary in place and carefully remove by gently pulling. Once free, put to one side place in a measuring cylinder for cleaning along with the IPA capillary inlet pieces.
- Disconnect the gas line and the exhaust hose from the main interface housing. Using the size 3 hexagonal key, unscrew the four nuts that connect the interface to the metal block. Using the same hexagonal key, unscrew the four screws that attach the metal block to the main instrument housing to reveal the StepWave. Place both the interface and the metal block on a clean surface ready for cleaning.
- Using the StepWave removal tool, hook underneath the StepWave and apply pulling pressure until the StepWave is released and can be guided out of the instrument housing. Support the full length of the StepWave at all times and avoid touching any of the circuitry.
- Prepare a 50:50 LCMS grade Water/IPA cleaning solution and label as Cleaning Solution.
- Pour a small amount of Cleaning Solution into a measuring cylinder large enough to accommodate the StepWave. Carefully, slide it into a separate measuring cylinder, using the liquid as a buffer against any damage. Alternatively, a plastic beaker large enough the StepWave is also appropriate.
- Small parts, such as the long and short capillaries and the Venturi sample inlet can also be placed in a beaker with some cleaning solution.
- Place in an ultrasound bath and sonicate for between 5-10 minutes. If the parts are particularly dirty, the sonication time can be extended to 20-30 minutes. The whistle in the Venturi interface and the kanthal coil should NOT be sonicated.
- Whilst the items are in the ultrasound bath, pour roughly 5ml of IPA into a small beaker. Wet the cotton tip of a wooden applicator with IPA and carefully wipe down all the surfaces on the metal block, the Venturi and the interface housing that have been exposed to samples, including at the capillary inlets and around the coil.
- The PEEK coil mounting plate can be removed from the main housing (take care as this is still attached by wires) and the internal surfaces of the interface housing can also be swabbed clean.
- When sonication is complete, gently empty the solution from the vessels into a beaker and leave all of the cleaned items on a paper towel to air dry. Switch on the gas flow from the microregulator and blow the StepWave dry to ensure there are no wet spots.
- Once all parts are dry, the elements can be reassembled.
- Before re-inserting the StepWave, check that the pins inside the main instrument housing are correctly orientated. Insert the StepWave into the instrument housing, supporting the full length as you do so. Gently push in until the StepWave clicks into place (there will be some degree of resistance).
- When the StepWave is in place, reposition the metal block and tighten the screws by degrees until it is sitting flush with the instrument housing.
- Supporting the weight, reposition the interface against the metal block and tighten the screws by degrees until it is fixed.
- Insert the capillary into the interface housing and push until the ferrule fixes it at

the correct distance from the coil. Secure in place using a nut.

- To pump down, go to MassLynx software → MS Tune (on vertical menu bar) → Vacuum → Pump down. This will activate the pumps in the instruments. Extend the view in MS tune so that the Vacuum tab is visible by going to View → Vacuum.
- Wait for approximately 3 minutes for the three turbo speeds to reach to 100% before leaving the instrument to reach the acceptable TOF pressure of less than 1.2×10^{-6} and backing pressure of between 1.2 and 2mbar before use (may take up to 4 hours). Placing a septum over the exposed end of the capillary will reduce the time taken to reach the required vacuum.

Troubleshooting

Lack of signal/poor signal quality

1. Check coil is on (Cmd Prompt 3024 or plugged into metal block below interface).
2. Check IPA is pumping (0.15 - 0.2ml/min).
3. Check air/nitrogen is on (2 bar on microregulator).
4. Check the right IPA capillary is being used (longer for calibration, shorter for tissue burning).
5. Check there is suction and all parts including gas are properly attached with no leaks.
6. Turn up the generator to a higher wattage if tissue is not burning properly.
7. Turn down the generator to a lower wattage if the spectrum is very noisy.
8. Check backing pressure and ToF pressure on vacuum tab. If backing pressure has changed, the capillary may be blocked. Try step 9 or 10.
9. Put through some methanol into instrument by very carefully touching a syringe with methanol quickly to the end of the MS capillary in case it might be blocked. Check signal and backing pressure to see if it has improved. DO NOT allow more than a few drops of solvent to go into the capillary as it will make the vacuum decrease (and it will vent if a lot of solvent is aspirated in). Check the ToF dial to ensure this hasn't happened.
10. Remove MS capillary and rinse through (water then IPA or methanol. Clean inside with thin capillary if available).

REM connection error on generator to return electrode

11. Unplug small blue plug from the generator, rotate 180° and plug in again.
12. Press down and smooth the electrode out.
13. Peel the electrode off carefully and re-stick it, smoothing outwards to ensure maximum contact area.
14. Peel off carefully, put a tiny amount of water on the black mat and wipe it across the area so it is very slightly damp but with no visible water then re-stick.
15. Cut more off the electrode, cutting evenly from each side around the electrode to retain symmetry around wire

Complete loss of signal, including electronic background signal or loss of communications

16. Switch to positive mode and back to negative mode.
17. Turn source into standby mode (yellow square) for 30 seconds and back to operate mode. Repeat several times if necessary, leaving between each time to check signal improvement.
18. Restart electronics: Put into pump override mode (top switch on side of

instrument) and switch off electronics (bottom switch) for ~10 seconds. Switch on electronics again and restart MassLynx.

19. Restart the computer and switch off electronics: Then use normal starting sequence -switch on electronics once computer has logged in, wait 5 minutes and open MassLynx.

20. Repeat the electronics restart several times with increasing length of electronics off (up to 5 mins), so as to maintain the vacuum before resorting to a full switch off (22).

21. Complete restart (loss of vacuum): Restart computer and turn off by unplugging and switching off electronics (as if going to surgery). Leave unplugged for 1 min and plug in and restart as normal.

22. If there is still no communication to the instrument, repeat the complete restart and leave the instrument unplugged for 30 mins. As a final step before contacting Waters, leave the instrument switched on overnight but not pumping and start MassLynx the following day.

ToF/pressure dials freezing

23. If ToF dial stays stuck at same vacuum level whilst pumping down (not full vacuum), restart electronics (18).

24. If restarting electronics fails to solve the problem, detach and reattach the cable to the Penning gauge (big red cylinder attached to ToF) inside the instrument side panel. Repeat several times

25. If this still fails to fix it, after detaching the cable, remove the head by twisting anticlockwise and pulling it off then gently tap the gauge with a spanner or similar item. If this also fails, a new gauge is needed – contact Waters.